

STIC-Biotech/ChemLib

93116

From: Rao, Manjunath N.  
Sent: Monday, May 05, 2003 8:52 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 09/868,328

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 6A04  
703-308-3534

RECEIVED  
MAY -5 2003  
STIC/CHEN, D. (STIC)

Date: 5-5-03

Please search the following as soon as possible for application with serial number **09/868328**

1. SEQ ID NO: 2 and 3 against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.
2. SEQ ID NO:1, 4 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO.

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/6  
Date Completed: 5/12  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:32 ; Search time 35.0673 Seconds  
(without alignments)  
1979.723 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845

Sequence: 1 MTPAISRAVLQAGAGALA.....TDGPAHFTGIVREIQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS2/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | ID          | Description        |
|------------|-------------|-------|--------|-------------|--------------------|
| 1          | 2845        | 100.0 | 521    | 22 AAB82301 | Arthrobacter ureaf |
| 2          | 2355        | 82.8  | 517    | 20 AAY04105 | Arthrobacter nicot |
| 3          | 2355        | 82.8  | 517    | 23 ABB10074 | Levan fructotransf |
| 4          | 2268        | 79.7  | 484    | 20 AAY04104 | Arthrobacter nicot |
| 5          | 2268        | 79.7  | 485    | 23 ABB10073 | Levan fructotransf |
| 6          | 499.5       | 17.6  | 943    | 21 AAY91929 | Paenibacillus mace |
| 7          | 468.5       | 16.5  | 1277   | 21 AAY91928 | Paenibacillus pabu |
| 8          | 461         | 16.2  | 923    | 21 AAY91927 | Paenibacillus amyl |
| 9          | 427         | 15.0  | 516    | 21 AAU76758 | Novel recombinant  |
| 10         | 416.5       | 14.6  | 533    | 19 AAU44863 | C. utilis INV1 pro |

|    |       |      |      |             |                    |
|----|-------|------|------|-------------|--------------------|
| 11 | 389.5 | 13.7 | 490  | 16 AAR72566 | Penicillium purpur |
| 12 | 386   | 13.6 | 750  | 20 AAY25346 | Bacillus sp. L7 en |
| 13 | 386   | 13.6 | 750  | 20 AAY24924 | Bacillus L7 endo-1 |
| 14 | 366   | 12.9 | 556  | 15 AAR56256 | Inulinase preprote |
| 15 | 359.5 | 12.6 | 532  | 5 AAP40020  | Saccharomyces cere |
| 16 | 359.5 | 12.6 | 532  | 12 AAR12145 | Preinvertase. Sac  |
| 17 | 348   | 12.2 | 581  | 20 AAY17500 | Schizosaccharomyce |
| 18 | 348   | 12.2 | 581  | 20 AAY08399 | S. pombe invertase |
| 19 | 348   | 12.2 | 581  | 21 AAB03200 | Schizosaccharomyce |
| 20 | 292.5 | 10.3 | 671  | 23 AAU74396 | Fructosyl transfer |
| 21 | 277.5 | 9.8  | 636  | 16 AAR75187 | Tomato plant inver |
| 22 | 277.5 | 9.8  | 636  | 21 AAB21098 | Wild-type tomato 1 |
| 23 | 274.5 | 9.6  | 636  | 13 AAR27007 | Tomato vacuolar in |
| 24 | 274.5 | 9.6  | 636  | 14 AAR35102 | Tomato acid invert |
| 25 | 266   | 9.3  | 583  | 22 AAE04416 | Lycopersicon penne |
| 26 | 266   | 9.3  | 648  | 23 AAU74395 | Fructosyl transfer |
| 27 | 261.5 | 9.2  | 1116 | 20 AAY43179 | S. rochei strain E |
| 28 | 258.5 | 9.1  | 645  | 23 ABB77844 | Amino acid sequenc |
| 29 | 244.5 | 8.6  | 479  | 23 ABB27213 | Streptococcus poly |
| 30 | 243   | 8.5  | 670  | 21 AAY85665 | Corn invertase pro |
| 31 | 241   | 8.5  | 553  | 15 AAR53922 | Tomato acid invert |
| 32 | 241   | 8.5  | 553  | 16 AAR82848 | Tomato acid invert |
| 33 | 239.5 | 8.4  | 1487 | 16 AAR74634 | Bacillus circulans |
| 34 | 235.5 | 8.3  | 492  | 23 ABB39050 | Staphylococcus epi |
| 35 | 235.5 | 8.3  | 591  | 21 AAG21596 | Arabidopsis thalia |
| 36 | 231.5 | 8.1  | 587  | 23 ABB92634 | Herbicidally activ |
| 37 | 223.5 | 7.9  | 569  | 23 ABB92300 | Herbicidally activ |
| 38 | 222.5 | 7.8  | 554  | 23 AAU74399 | Fructosyl transfer |
| 39 | 222   | 7.8  | 591  | 23 ABB91362 | Herbicidally activ |
| 40 | 220   | 7.7  | 581  | 23 ABB92301 | Herbicidally activ |
| 41 | 220   | 7.7  | 588  | 21 AAG50247 | Arabidopsis thalia |
| 42 | 205.5 | 7.2  | 630  | 17 AAU00455 | l-Sucrose:sucrose  |
| 43 | 205.5 | 7.2  | 637  | 19 AAW77034 | Artichoke sucrose  |
| 44 | 202.5 | 7.1  | 429  | 22 AAB83190 | Corynebacterium th |
| 45 | 202.5 | 7.1  | 433  | 22 AAG92651 | C glutamicum prote |

ALIGNMENTS

RESULT 1  
AAB82301  
ID AAB82301 standard; Protein: 521 AA.  
AC AAB82301;  
XX  
XX  
DT 09-JUL-2001 (first entry)  
XX  
DE Arthrobacter ureafaciens levan fructotransferase.  
XX  
KW Levan fructotransferase; difructose dianhydride IV; sweetener.  
XX  
OS Arthrobacter ureafaciens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..33  
FT /label= Signal\_peptide  
FT Protein 34..521  
FT /label= Mature\_protein  
XX  
PN WO200129185-A1.  
XX  
PD 26-APR-2001.  
XX  
PE 19-OCT-2000; 2000WO-KR01183.  
XX  
PR 19-OCT-1999; 99KR-0045302.  
XX  
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.  
PA (REAL-) REALBIOTECH LTD.  
XX  
PI Rhee S, Song K, Kim C, Ryu E, Lee Y;  
XX

DR WPI; 2001-308483/32.  
 XX N-PSDB; AAF30918, AAF30919.  
 PT Producing difructose dianhydride IV from sucrose, involves reacting  
 PT sugar solution in the presence of levansucrase to produce levan, and  
 PT reacting levan solution in the presence of levan fructotransferase to  
 PT produce DFA IV -  
 XX  
 PS Claim 2; Page 47; 72pp; English.  
 XX  
 CC The present sequence is that of *Arthrobacter ureafaciens* K2032  
 CC levan fructotransferase (see AAB82301). The enzyme can be  
 CC obtained by cultivation of claimed *Escherichia coli* JUD81  
 CC (KCTC 0877BP), which carries claimed expression vector pUDFA81  
 CC comprising the levan fructotransferase gene (see AAF30918). A  
 CC claimed process for producing difructose dianhydride IV from  
 CC sucrose comprises subjecting sugar solution to reaction at room  
 CC temperature or lower in acidic buffer of pH 3.0-7.0 in the presence  
 CC of a levansucrase derived from *Zymomonas mobilis* to produce levan,  
 CC purifying the levan from the reaction solution, and subjecting it  
 CC to reaction at 25-30 degree C for 3-10 hours in acidic buffer of pH  
 CC 3.0-7.0 in the presence of levan fructotransferase, preferably  
 CC obtained from *E. coli* JUD81. The product is useful as a low-calorie  
 CC sweetener.  
 XX  
 PS Sequence 521 AA;

Query Match 100.0%; Score 2845; DB 22; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-241;  
 Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTPAISRRVLOGAGAGALALIFGGVPPAARASAPGSLRAVYHMTPPSGWLCDDPQPVY 60  
 DB 1 MTPAISRRVLOGAGAGALALIFGGVPPAARASAPGSLRAVYHMTPPSGWLCDDPQPVY 60

QY 61 THGAYQLYLHSDQNGPGGDHASTTGDGVAFTHHGTGTMPLRDPFVWSSAVVGTA 120  
 DB 61 THGAYQLYLHSDQNGPGGDHASTTGDGVAFTHHGTGTMPLRDPFVWSSAVVGTA 120

QY 121 GFGAGAVVALATOPTDGVKRYQYLYWSTGDTGFTTALPDVIVNTDGRAATPAEIE 180  
 DB 121 GFGAGAVVALATOPTDGVKRYQYLYWSTGDTGFTTALPDVIVNTDGRAATPAEIE 180

QY 181 AEWFRDPKTHWDTARGWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
 DB 181 AEWFRDPKTHWDTARGWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240

QY 241 FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTGWDGQFHADDLTPOWLDGWWDYAA 300  
 DB 241 FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTGWDGQFHADDLTPOWLDGWWDYAA 300

QY 301 VTWPSIDAPETKRLATAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360  
 DB 301 VTWPSIDAPETKRLATAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360

QY 361 STPVAALTNYVTATTTLPDRVDSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420  
 DB 361 STPVAALTNYVTATTTLPDRVDSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420

QY 421 RHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSVEVFNA 480  
 DB 421 RHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSVEVFNA 480

QY 481 GHTVLSQQVHFAEGDGTGISLYTDGGPAHFTGIVVREIGQAI 521  
 DB 481 GHTVLSQQVHFAEGDGTGISLYTDGGPAHFTGIVVREIGQAI 521

RESULT 2  
 AAY04105  
 ID AAY04105 standard; Protein; 517 AA.  
 XX  
 AC AAY04105;

XX 10-JUN-1999 (first entry)  
 XX  
 DE Arthrobacter nicotinovorans levan fructotransferase protein #2.  
 KW Arthrobacter nicotinovorans; levan fructotransferase.  
 XX  
 OS Arthrobacter nicotinovorans.  
 PN JP11069978-A.  
 XX  
 PD 16-MAR-1999.  
 XX  
 PF 28-AUG-1997; 97JP-0232421.  
 PR 28-AUG-1997; 97JP-0232421.  
 XX (NIOC) NIPPON OIL CO LTD.  
 PA WPI; 1999-247463/21.  
 DR N-PSDB; AAX19827.  
 XX  
 PT Levan fructotransferase gene - for recombinant production of levan  
 PT fructotransferase  
 XX  
 PS Claim 3; Page 8-9; 14pp; Japanese.  
 CC The present sequence represents *Arthrobacter nicotinovorans* levan  
 CC fructotransferase. The present invention also describes a method  
 CC for the preparation of levan fructotransferase in which a transformant  
 CC is cultured in a medium and levan fructotransferase is collected from  
 CC the culture. The method can prepare levan fructotransferase in a  
 CC large amount.  
 XX  
 PS Sequence 517 AA;

Query Match 82.8%; Score 2355; DB 20; Length 517;  
 Best Local Similarity 81.0%; Pred. No. 7.1e-198;  
 Matches 422; Conservative 41; Mismatches 54; Indels 4; Gaps 2;

QY 1 MTPAISRRVLOGAGAGALALIFGGVPPAARASAPGSLRAVYHMTPPSGWLCDDPQPVY 60  
 DB 1 MTYDISRRALOGAGAGALALFMSNAIPVAHAQA--SLRAIYHMTPPSGWLCDDPQPVH 58

QY 61 THGAYQLYLHSDQNGPGGDHASTTGDGVAFTHHGTGTMPLRDPFVWSSAVVGTA 120  
 DB 59 TNGAYQLYLHSDQNGPGGDHASTTGDGVSYTHGHVWMPQDPFVWSSAVVDTA 118

QY 121 GFGAGAVVALATOPTDGVKRYQYLYWSTGDTGFTTALPDVIVNTDGRAATPAEIE 180  
 DB 119 GFGAGAVIALATOPTDG--KFQEQYLYWSTGDTGYSFTALPDVIVNTDGRATTA 176

QY 181 AEWFRDPKTHWDTARGWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
 DB 177 AEWFRDPKTHWDATRNSEWCVIGRLRYAAYFTSPNLRDQWQNSNFDYPNHALGGIECPDL 236

QY 241 FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTGWDGQFHADDLTPOWLDGWWDYAA 300  
 DB 237 FEMTAGDGTTHWVFGASMDAYSIGLPMYAYWTGWSNGTAFADNLTPOWLDGWWDYAA 296

QY 301 VTWPSIDAPETKRLATAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360  
 DB 297 VTWPAVEAPETKRLATAMNNWKYAAARNVPTDASDGYNGQNSITRELRLERQSGWYTL 356

QY 361 STPVAALTNYVTATTTLPDRVDSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420  
 DB 357 STPVPALSNYATSSITLPDRVTNGSEVLWPSGRAYEIELDISWDTAANVCVGRSSDGS 416

QY 421 RHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSVEVFNA 480  
 DB 417 RHTNICKYGDELIVDRASSEQSGYALAPYTRAAAPIDANARSVHLRIFVDQTSVEVFNS 476

QY 481 GHTVLSQQVHFAEGDGTGISLYTDGGPAHFTGIVVREIGQAI 521



QY 278 GEFHADDLPQWLDGMDWYAAVTPSIDAPETKRLAIAMNNWYAAARDVPTDASDGY 337  
 DB 241 GTAFIADNLTPQWLDGMDWYAAVTPPAVEAPETKRLATAMNNWYAAARNVPTDASDGY 300  
 QY 338 NGNSIVRELRLAROPGGWYLLSTPVAALNTNYVATTTLPLDRTVDSGLPWNHGRAYEI 397  
 DB 301 NGNSITRELRLEROSGGWYLLSTPVPALSNYATSTTLPLDRTVNGSFVLPWNGRAYEL 360  
 QY 398 ELDIAMDTATNVGISVGRSDGSRHTNIGYGDADLYVDRGPSDLAGYSLAPYSRAAAPID 457  
 DB 361 ELDISMDTAANVGVSGRSDGSRHTNIGYGDADLYVDRASSQSGYALAPYTRAAAPID 420  
 QY 458 PGARSVHLRLILVDTQSVVEFVNAGHTVLSQOVHFAEGDTGISLYTDGPPAHFTGIIVVREI 517  
 DB 421 ANARSVHLRFVDTQSVVEFVNSGHTVWSQOVHFAAGDTGISLYADGGPANFTGITIREF 480  
 QY 518 GOAI 521  
 DB 481 GNPI 484

RESULT 5  
 ABB10073  
 ID ABB10073 standard; protein; 485 AA.  
 AC ABB10073;  
 DT 01-JUL-2002 (first entry)  
 XX Levan fructotransferase amino acid sequence.  
 DE Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;  
 KW Levan; enzyme.  
 OS Arthrobacter nicotinovorans.  
 XX  
 PN JP2002017366-A.  
 XX  
 PD 22-JAN-2002.  
 XX  
 PF 06-JUL-2000; 2000JP-0205756.  
 XX  
 PR 06-JUL-2000; 2000JP-0205756.  
 XX  
 PA (NIPT ) NIPPON TENSAI SEITO KK.  
 PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 XX  
 DR WPI; 2002-287313/33.  
 DR N-PSDB; ABB10073.  
 XX  
 PT A new levan fructotransferase -  
 XX  
 PS Claim 2; Fig 1; 15pp; Japanese.  
 XX  
 CC The invention relates to a new levan fructotransferase (LFTase). LFTase  
 CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to  
 CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention  
 CC is used for the preparation of DFA IV. The current sequence represents  
 CC the levan fructotransferase amino acid sequence.  
 XX  
 SQ Sequence 485 AA;

Query Match 79.7%; Score 2268; DB 23; Length 485;  
 Best Local Similarity 82.6%; Pred. No. 2.8e-190;  
 Matches 400; Conservative 40; Mismatches 42; Indels 2; Gaps 1;

QY 38 SLRAVYHMTPPSGWLCDDPQPVTHGAYQLYLLHSDQNGPGGWDHASTDGVAFTHHGT 97  
 DB 4 SLRAIYHMTPPSGWLCDDPQPVTHGAYQLYLLHSDQNGPGGWDHATTGDDGVSYTHHGV 63  
 QY 98 VMLPRDFPWSGSAVVGTANTAGFCAGAVVALATOPTDGVKRYQSOXYLWSTDGGFTET 157  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||

Db 64 VMPQPDFFPWSGSAVVDTANTAGFCAGAVIALATOPTDG--KFQEQYLYWSTDGGYSFT 121  
 QY 158 ALPDPVIVNTDGRAATTTPAEIENAEWFRDPKIHWDTARGEWCVIGRLRYAAAYTSPNLR 217  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
 Db 122 ALPDPVIVNTDGRATTTPAEVENAEWFRDPKIHWDATRNWCVIGRARYAAAYTSPNLR 181  
 QY 218 DWTLRNFDYPNHALGGIECPDLFEITADGTRHWWLAASMDAYGICLPMYAYWTGTWD 277  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
 Db 182 DMQKSNFDYPNHALGGIECPDLFEMTAGDGTTRHWFEGASMDAYSICLPMFTFAYTGSWN 241  
 QY 278 GEFHADDLPQWLDGMDWYAAVTPSIDAPETKRLAIAMNNWYAAARDVPTDASDGY 337  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
 Db 242 GTAFIADNLTPQWLDGMDWYAAVTPPAVEAPETKRLATAMNNWYAAARNVPTDASDGY 301  
 QY 338 NGNSIVRELRLAROPGGWYLLSTPVAALNTNYVATTTLPLDRTVDSGLPWNHGRAYEI 397  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
 Db 302 NGNSITRELRLEROSGGWYLLSTPVPALSNYATSTTLPLDRTVNGSFVLPWNGRAYEL 361  
 QY 398 ELDIAMDTATNVGISVGRSDGSRHTNIGYGDADLYVDRGPSDLAGYSLAPYSRAAAPID 457  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
 Db 362 ELDISMDTAANVGVSGRSDGSRHTNIGYGDADLYVDRASSQSGYALAPYTRAAAPID 421  
 QY 458 PGARSVHLRLILVDTQSVVEFVNAGHTVLSQOVHFAEGDTGISLYTDGPPAHFTGIIVVREI 517  
 |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
 Db 422 ANARSVHLRFVDTQSVVEFVNSGHTVWSQOVHFAAGDTGISLYADGGPANFTGITIREF 481  
 QY 518 GOAI 521  
 Db 482 GNPI 485

RESULT 6  
 AAY91929  
 ID AAY91929 standard; Protein; 943 AA.  
 AC AAY91929;  
 XX  
 DT 19-JUL-2000 (first entry)  
 XX  
 DE Paenibacillus macerans 2,6-beta-D-fructan hydrolase.  
 XX  
 KW 2,6-beta-D-fructan hydrolase; anticarie; neuroprotective; biofilm;  
 KW hydrolyze slime; detergent; fructose production; sweetener.  
 XX  
 OS Paenibacillus macerans.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 29..943  
 FT /label= mature\_protein  
 XX  
 PN WO200017331-A1.  
 XX  
 PD 30-MAR-2000.  
 XX  
 PF 17-SEP-1999; 99WO-DK00495.  
 XX  
 PR 18-SEP-1998; 98DK-0001173.  
 PR 09-DEC-1998; 98DK-0001623.  
 XX  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Moller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;  
 XX  
 DR WPI; 2000-283564/24.  
 DR N-PSDB; AAA08535.  
 XX  
 PT New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful  
 PT e.g. for degrading microbial biofilm, hydrolyzing slime, production of  
 XX fructose oligomers and treatment of dental plaque  
 PS Claim 1; Page 119-120; 123pp; English.  
 XX  
 CC AAY91927-29 are polypeptides from Paenibacillus amylolyticus, P. papuli

CC and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase  
CC activity. These are preferred enzymes of the invention. The enzymes  
CC can be used to remove microbial biofilm from surfaces, particularly for  
CC reduction of bioerosion, for treatment of dental plaque and for  
CC treatment of multiple sclerosis (eliminating biofilm from the lungs; to  
CC hydrolyze slime; in detergents and for production of fructose oligomers  
CC from levan or phleix (useful as sweeteners or for conversion to  
CC fructose). The nucleic acids that encode the enzymes are useful for  
CC recombinant production and as source of probes and primers for  
CC identification and cloning of related fructan hydrolases (all claimed).  
XX  
XX  
SQ Sequence 943 AA;

Query Match 17.6%; Score 499.5; DB 21; Length 943;  
Best Local Similarity 30.8%; Pred. No. 1.2e-34;  
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;

QY 40 RAYVHTPPSGWLCDDPQPVTHGAYQVLYLHSDQNGGWDHSTTDCVAFTHGTVM 99  
DB 383 RPDYHYPARGASADPNGLVYEGEYHLF--HQD---GSTWAHAYSTDLVHWKRLPIAL 436  
QY 100 PLRPDPFVWVSGSAVGTANTAGF-----GAGAVVALATOPTGVRKYQCYLYWSTDGGF 154  
DB 437 PWNLDGHVWSSAVADLHNASGLFADSGKGLIAYTTSYNPDRPQNGRIGLAYSKDRGR 496  
QY 155 TTF-ALPDPVIVNTDGRAATTPAEIENAEW-FRDPKIHMDTARGEVWCVIGRLRYAAFTY 212  
DB 497 TWEYAAERPIVIEPNKQGGDDP-----GGWDFRDPKVRDEEHNWVWVSGGDHIFFT 551  
QY 213 SPNLRDWTLLRNFDYNHNLGGI-ECDFLFEITADD-GTRHWVLAASMDAYGIGLPMTYA 270  
DB 552 STNLIDTTLTDSFGYAYVGVWGPCDLFLQLAVDVTGKWKVLMISTGAN-----PNTQG 607  
QY 271 YMTGTWDGEGFHADLTTPQW-----LDGWDWYAAVWTPSIDAPETKRLATA 317  
DB 608 -----SAAEYFGEITPEKGFNNDPNACKVLTADTGKEYYASMSFAGM--PDGRRVMLA 659  
QY 318 WMNNKYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLSTPVAALTN----YVTA 373  
DB 660 WMTNWDYPPA-FPT---EGWKGVLSIPRELTQKTDKG-IRLAQTPIRELESURGQLLFA 714  
QY 374 TITLPTDRTVDG---SAVLPNWGRAYEIELDIANDTATNNGISVGRSPDGRTRHNTICKYGA 430  
DB 715 AS---DRRVOADRENILKGVSSGVYIEAEIETPQASVNSEFGRLREGAKRTVVGYKT 771  
QY 431 ---DLVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLTVDTSQVSEVFNAGHTVLSQ 487  
DB 772 KENEIVDRSLSGDTGFSERFTLHOAPLOPDNRNRVKKLRFVDDSSLEFVGDDGRVVFSE 831  
QY 488 QVHFAEGDTGISLYTDGPAHFTGIVVREI 517  
DB 832 VIFPPDPAHREMSLFTVGGEVNVSLLKVALH 861

RESULT 7  
AAAY91928  
ID AAY91928 standard; Protein; 1277 AA.  
AC AAY91928;  
XX  
XX 19-JUL-2000 (first entry)  
XX Paenibacillus pabuli 2,6-beta-D-fructan hydrolase.  
XX 2,6-beta-D-fructan hydrolase; antidiabetic; neuroprotective; biofilm;  
KW hydrolyze slime; detergent; fructose production; sweetener.  
XX  
XX Paenibacillus pabuli.  
XX  
XX Key Location/Qualifiers  
FH 25..1277  
FT Protein /label= mature\_protein  
XX

PN WO200017331-A1.  
XX 30-MAR-2000.  
XX 17-SEP-1999; 99WO-DK00495.  
XX 18-SEP-1998; 98DK-0001173.  
PR 09-DEC-1998; 98DK-0001623.  
XX (NOVO ) NOVO-NORDISK AS.  
XX  
XX Moller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;  
PI WPI; 2000-283564/24.  
DR N-PSDB; AAA08533.  
XX New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful  
PT e.g. for degrading microbial biofilm, hydrolyzing slime, production of  
PT fructose oligomers and treatment of dental plaque  
XX  
PS Claim 1; Page 111-115; 123pp; English.  
XX  
CC AAY91927-29 are polypeptides from Paenibacillus amylolyticus, P. pabuli  
CC and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase  
CC activity. These are preferred enzymes of the invention. The enzymes  
CC can be used to remove microbial biofilm from surfaces, particularly for  
CC reduction of bioerosion, for treatment of dental plaque and for  
CC treatment of multiple sclerosis (eliminating biofilm from the lungs; to  
CC hydrolyze slime; in detergents and for production of fructose oligomers  
CC from levan or phleix (useful as sweeteners or for conversion to  
CC fructose). The nucleic acids that encode the enzymes are useful for  
CC recombinant production and as source of probes and primers for  
CC identification and cloning of related fructan hydrolases (all claimed).  
XX  
SQ Sequence 1277 AA;

Query Match 16.5%; Score 468.5; DB 21; Length 1277;  
Best Local Similarity 29.1%; Pred. No. 1e-31;  
Matches 168; Conservative 74; Mismatches 243; Indels 93; Gaps 25;

QY 4 AISRRAVLOGAGA-----GALALIFGAVPPAA-----RASAPG----- 37  
DB 320 ASSNRKVPQSGARHHEIVASGLIQVYVDGYTPAAVEVTDKSYAKGNAGLVVOOGMAY 379  
QY 38 -----SLRAYVHTPPSGWLCDDPQPVTHGAYQVLYLHSDQNGGWD 82  
DB 380 FQDIYMTESMYKENYRQYHYSPLRGSADPNGLVYEGEYHLF--HQD---CGTWA 433  
QY 83 HASTTDGVAFTHTGTVMLRPPDPFVWVSGSAVGTANTAGF-----GAGAVVALATOPTDG 137  
DB 434 HAVSSDLINWKRPLIALPNWDGHWVSGSAIADLNNASGLFTDSDGKGLIAYTSYHPDK 493  
QY 138 VRKYQEOYLYWSTDGGETFT-ALPDPVIVNTDGRAATTPAEIENAEW-FRDPKIHMDTAR 195  
DB 494 PGQNQRIGLAYSTDQGRNQYAKERIVINDPNKNGDDP-----GSWDFRDPKVVREDEH 548  
QY 196 GEWCVIGRLRYAAFTSPNLRDWTLLRNFDYNHNLGGI-ECDFLFEITAD-DGPRHW 253  
DB 549 NRWVWVSGGDHIFRTSTNLLDNLTLTDFNGYDYGVRGSGWGPCDLIQLPVDGTGQKKV 608  
QY 254 LAASDAYGIGLPMTYA----YWTG--TWDEGEFHADLTTPQWL--DWGWDYAAVWTPS 305  
DB 609 LLISTGAN---PKTQSDAEYFVGQLTADG-KFLNDHPAGQVLRDTCYKEFYASNSFAN 663  
QY 306 IDAPETKRLAIAMNNKYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLSTPVA 365  
DB 664 M--PNQKRVNLAAMTNDYPPF-EFTSS---WKGQLTIPREVSRLRTTDEG-VRLVQTPIT 716  
QY 366 ALTNVYVATTTLPDRTVDCSAVLPMNG---RAYEIELDI---AWDTATNNGISVGRSPDG 419  
DB 717 ELQKLRHNLVSAQQMTVGPKNPGLTAGAYEIEAEVEIPANSVTEFGFOL-KOREG 775  
QY 420 TRHTNIGKYGAD---LYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLTVDTSQVSEV 476

Db 776 QKTT--VAYRVDTQNMVFDRTTSGVFSFSDLFTKVHEASLKPENQKVKLRIFVDESSVEV 833  
Qy 477 FVNGHTVLSQOVHFAEGDGTGSLYTDGGPAHFTGIIV 514  
Db 834 FGDKGVVFSVDVIFPDAGRAMAFYSLGGEVKVSSMKV 871

RESULT 8  
AAV91927  
ID AAV91927 standard; Protein; 923 AA.  
AC AAV91927  
XX 19-JUL-2000 (first entry)  
DE Paenibacillus amylolyticus 2,6-beta-D-fructan hydrolase.  
XX 2,6-beta-D-fructan hydrolase; anticaric; neuroprotective; biofilm;  
KW hydrolyze slime; detergent; fructose production; sweetener.  
XX Paenibacillus amylolyticus.  
XX Key Location/Qualifiers  
FT Protein 32..923  
FT /label= mature\_protein  
XX WO200017331-A1.  
XX 30-MAR-2000.  
XX 17-SEP-1999; 99WO-DK00495.  
XX 18-SEP-1998; 98DK-0001173.  
XX 09-DEC-1998; 98DK-0001623.  
XX (NOVO ) NOVO-NORDISK AS.  
XX Moller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;  
XX WPI: 2000-283564/24.  
XX N-PSDB; AAA08533.  
XX New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful  
PT e.g. for degrading microbial biofilm, hydrolyzing slime, production of  
PT fructose oligomers and treatment of dental plaque  
XX Claim 1; Page 107-110; 123pp; English.  
XX AAV91927-29 are polypeptides from Paenibacillus amylolyticus, P. pabuli  
CC and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase  
CC activity. These are preferred enzymes of the invention. The enzymes  
CC can be used to remove microbial biofilm from surfaces, particularly for  
CC reduction of biocorrosion, for treatment of dental plaque and for  
CC treatment of multiple sclerosis (eliminating biofilm from the lungs; to  
CC hydrolyze slime; in detergents and for production of fructose oligomers  
CC from levan or phlelin (useful as sweeteners or for conversion to  
CC fructose). The nucleic acids that encode the enzymes are useful for  
CC recombinant production and as source of probes and primers for  
CC identification and cloning of related fructan hydrolases (all claimed).  
XX Sequence 923 AA;  
Query Match 16.2%; Score 461; DB 21; Length 923;  
Best Local Similarity 29.5%; Pred. No. 2.8e-31;  
Matches 150; Conservative 74; Mismatches 223; Indels 62; Gaps 23;

Qy 40 RAYVHTPPSGWLCDDQRPVTHGAYQLYLHSDQNNPGCGWDHASTDCGVAETHHGTVM 99  
Db 392 RPQYHTPIRGSASDPNGLYFEGEYHLF--HQD----GGTWAHVAVSKDMLNKRPLIAL 445  
Qy 100 PLRPDPFVWGSASVVGTAANTAGF-----GAGAVVALATQPTDGVKRYOEQLYWSIDGGF 154

Db 446 PWNHGHVWGSASVADMTNAGSLFGSGGKGLIAYTTSFNPDPSPNGNORIGLAYSKDQGR 505  
Qy 155 TFTAALPD-PVIVNTDGRAATTPAEIEAENAEW-FRDPKTHWDTARGEWVCVIGRLRYAAFYT 212  
Db 506 TWEYSKERPIVENPGKSGN-----EAGNDFRDPKVIIRDENNRRVWVWVSGGDHIREYT 560  
Qy 213 SPNLRDWTLLRRNFDYPNHALGGI-ECPDLFEITADGCT--RHVYLAASMDAYGIGLPMY- 268  
Db 561 STNLLDWTLLDNMGYGDYVRGGVWECPLQLPV-DGTSQKWKVMMISTGAN-----PKTG 615  
Qy 269 ---YAYWTG--TWDEGOFHADDLTPOWL--DWGWDWAAVWTPSIDAPETKRLAIAMNN 321  
Db 616 GSDAEYFIGHLTAG-KFVNDNPAGKVLRTDFGKEFYASMFANM--PDHRTVMMAWMTN 672  
Qy 322 WKYAARDVPTDASDYGNGQNSIVRELRLARQPGGWYLLSTFPVAAALNYVTATTTLPDRT 381  
Db 673 WDYPFA-FPT---SNWKGELTIPREVSIVTTEDEG-IRMVQSPIKEL-----ESLRKPLYS 722  
Qy 382 VDSAVLPWNGR-----AYEIELDI---AMDATNVGISVGRSPDGTTRHNIKGYA 430  
Db 723 ASNKSVPSSGNLLKGIISGAYEIEAIEIPETSTVTEFGFNIREG--ANQKTWVGKAS 780  
Qy 431 D--LYVDRGPSDLAGYSLAPYSRAAAPIDFGARSVHLRLILVDTSQSVFVFNAGHTVLSQ 488  
Db 781 DSRMEVDRTASGETDFSNLFSKKHEAPTQMNENNRKIRLILVDESSVEAFGNDGKVFSDV 840  
Qy 489 VFHAEAGDTGISLYTDGGPAHFTGIIVREI 517  
Db 841 IFPDPAASRAMSFYKGGNVNVSILKVHQL 869

RESULT 9  
AAU76758  
ID AAU76758 standard; Protein; 516 AA.  
XX AAU76758;  
AC AAU76758;  
DT 20-MAY-2002 (first entry)  
XX Novel recombinant Aspergillus endoinulinase associated protein.  
DE Recombinant expression vector; Aspergillus endoinulinase;  
KW transformed microorganism.  
XX Unidentified.  
XX KR2000004834-A.  
XX 25-JAN-2000.  
XX 10-JUL-1998; 98KR-0027920.  
XX 10-JUN-1998; 98KR-0021432.  
XX (UHMT/) UHM T B.  
PA (CHAE/) CHAE G S.  
PA (HOIT/) HOIT J V.  
XX Uhm TB, Chae GS;  
XX WPI: 2000-677409/66.  
DR N-PSDB; ABK10615.  
XX Novel recombinant Aspergillus endoinulinase gene derived expression  
PT vector and transformed microorganism - NoAbstract  
XX Claim 2; Page 10; 12pp; Korean.  
XX The present invention relates to a new recombinant Aspergillus  
CC endoinulinase gene derived expression vector and also describes a  
CC transformed microorganism. The present amino acid sequence represents  
CC the novel recombinant Aspergillus endoinulinase gene derived expression  
CC vector protein of the invention.

```
XX SQ Sequence 516 AA;
Query Match 15.0%; Score 427; DB 21; Length 516;
Best Local Similarity 27.2%; Pred. No. 1.2e-28;
Matches 145; Conservative 92; Mismatches 221; Indels 76; Gaps 21;
Qy 24 GGAVPPAARASAGSLRAYVHTPPSGWLCDPQRPVTHGAYOLYVLSHSDNNGPGG--W 81
Db 16 GLTLPSQAQSN---DYRSPYHFTPDQYWNNEPGLIKIGTWHLFFQHNPNTANWGNICW 72
Qy 82 DHASTDGVATHTHCTVPLRPDPVWSGSAVVGTAANTAGFAGACA---VVALATOPTDGV 138
Db 73 GHATSTDLHMAHKPTALADENGVEAFTGTAYDPNNTSGLGDSANPPYLAFTGYTTS- 131
Qy 139 RYQEOYLWSTGDGFTTALPDVIVNTDGRAATTAEIENAEFRPKIHWDTARGEW 198
Db 132 SOTQORLAFSVNDGATWTKQGNPIIITSQEA---PHDITGGLESRDPKVFHFRQSNW 188
Qy 199 VCVI--GRLRYAAFTSPNLRDWTLLRRNFDYPNHALGGI-----ECPDLFEITADDGT 249
Db 189 IMVLAHGQDKLSEWTSADTINTWQS--DLKSTSIINGLSSDITGWEPDMFELPV-EGT 245
Qy 250 RH--WVL----AASMDAYIGLPMYIAYTGTWDEQFHAD--DLTPQWLQWGDWYAAV 301
Db 246 BETTWVNMTPAEGSPAGGNGVLAI---TGSFDCGKSFADPVDASTMWLQNGRDFDGL 301
Qy 302 TWPSIDAPETKRLATAWNNKYYAARDVPTDASDGYNGCONSTVRELRLAROPGGWYLLS 361
Db 302 SWNVNPSADGRRIIAVWNSY---GSPNPTTT---WKMLSPFPRLSL-KKVGTOOHRVQ 354
Qy 362 TPVAALTYVATTTLPTDRTYDGSVAVLPWNGRAYEIELDIAW--DTATNVGISVGRSPDG 419
Db 355 QPITELDTFISTSLQILANQITTPGQTLSSIRGTALDVRVAFYPDAGSVLSIAVRK---G 411
Qy 420 TRHNTIGKY---GADLYVDRGPSDLAGYSLAPYSRAAPIDPGARSVH-----464
Db 412 ASEQTVIKYQSDATLSVDRTESGDISY-----DPAAGGVHTAKLEEDGTGLV 459
Qy 465 -LRILVDPOSVEFVFNAGTIVLSQCVHFAEGDGTGISLTDGPGPAHFTGIVVREI 517
Db 460 SIRVLVDTCSEVVEVGQGEANISDLIFPSSDSDGLALEVTGNAVLSQVDVRSV 513

RESULT 10
AAW44863
ID AAW44863 standard; Protein; 533 AA.
XX AC AAW44863;
XX DT 29-SEP-1998 (first entry)
XX DE C. utilis INV1 protein.
XX KW URA3; HIS3; INV1; orotidine 5'-monophosphate decarboxylase; enzyme;
XX KW biosynthesis; ribonucleotide; uracil; histidine; auxotroph; invertase;
XX KW imidazole-glycerol-phosphate dehydratase; beta-fructofuranosidase;
XX KW genetic marker; plasmid; transformation.
XX OS Candida utilis.
XX FN WO9814600-A1.
XX PD 09-APR-1998.
XX PF 03-OCT-1997; 97WO-CU000005.
XX PR 03-OCT-1996; 96CU-00000082.
XX PA (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
XX PI Besabe Tuero L, Chavez Espinoza FP, Delgado Boada JM;
XX PI Gonzalez Martinez ME, Paifer Reyes E, Rivero Baeza T;
```

```
PI Rodriguez Menocal L;
XX WPI; 1998-240099/21..
XX N-PSDB; AAV19464.
XX PT New Candida utilis auxotrophic mutants and DNA markers - useful in
XX PT transformation systems to allow selection of transformants, e.g. for
XX PS heterologous protein production
XX PS Disclosure; Page 29-31; 63pp; Spanish.
XX CC This is the amino acid sequence of the Candida utilis INV1 anzyme
XX CC invertase (beta-fructofuranosidase). The invention relates to novel
XX CC C. utilis strains lacking one of the URA3 (AAV19462), HIS3 (AAV19463) or
XX CC INV1 genes resulting in auxotrophic mutants. The genes can be used
XX CC as markers in plasmids for transforming C. utilis. This organism has
XX CC not been used as a host for transformation due to difficulties in
XX CC obtaining reliable markers. Straightforward selection of transformants
XX CC using the new auxotrophic mutants as host cells is now possible.
XX SQ Sequence 533 AA;
Query Match 14.6%; Score 416.5; DB 19; Length 533;
Best Local Similarity 25.9%; Pred. No. 1e-27;
Matches 132; Conservative 87; Mismatches 182; Indels 109; Gaps 22;
Qy 40 RAVYHMTPPSGWLCDPQRPV--TTHGAYOLYVLSHDON---NGPGWDHASTTDCGVAFTH 94
Db 30 RPLVHLTPPGVGMNDPGLFYDSSSTHYVYQYVNPNTINGLPLYWGHATSDLLTWDH 89
Qy 95 HG--TYMPLRDPFPVWWSGSAVVGTAANTAGFAGACA---VVALATOPTDGVKRYQEOYL 147
Db 90 HAPAIQENDEGIVSGSIVDIDYDNTSGFDDSTPEQRIVAIYTNLPLDV---ETQDIA 146
Qy 148 WSTDGFTTALPDVIVNTDGRAATTAEIENAEFRPKIHWDTARGEWVCVIGRLR- 206
Db 147 YSTDGTYTFEYENNPVIDV-----NSTQFRDPKVIWYEETQOWMTYAKSOE 194
Qy 207 -YAAFTSPNLRDWTLLRRNFDYPNHALGIECPDLFEITAD-----DGTTRHWVLAASMDA 260
Db 195 YKIQIYTSNLDKWSLASNFSTKGVGYQYECPLFEATIENTPKSGDPEKKWVWLAINP 254
Qy 261 YGIGLPM---TYAYWTGTWDEQFHADDLTPQWLQWGDWYAAVTPWPSIDAPETKRLAIA 317
Db 255 ---GSPLGGSINEXFVGDFNGTEFTPDDDATRFMDTKDFYAFQAF--FNAPENRSIGVA 309
Qy 318 WNNNKKYAADVPTDASDGYNGCONSVIRE-----LRLARQPGGWYLLSTP 363
Db 310 WSSNMQY-SNQVPD--PDGYRSSMSSIREYTLRYVYSTNPESQLILCQK---FFVNETD 363
Qy 364 VAALTYVATTTLPTDRTVD---QSAVLPMNGRAYEIELDIAWDTATNVGISVGRSPDGT 420
Db 364 LKVEEYKVSNSL---TVDHFTGSSFANSNTTGL---LDP-----NMTFTVNGTTDVT 411
Qy 421 RHTNIGKYGADLYVDRGPSDLA---GYSL-----APYSR 451
Db 412 QKDSV---TEELRIKSNQSDAIALGYDYNNEQFYNRATESYFQRTNQFOERSTVVO 468
Qy 452 AAAPIDPGARSVHLRLIYDVTQSVEVFVNAV 481
Db 469 PLTITESGDKQYQLYGLVDNNILELYFNDG 498

RESULT 11
AAR72566
ID AAR72566 standard; Protein; 490 AA.
XX AC AAR72566;
XX DT 23-NOV-1995 (first entry)
XX DE Penicillium purpurogenum endo-inulinase.
XX XX
```

KW Penicillium purpurogenum variety rubrisclerotium; endo-inulinase;  
 KW inulo-oligosaccharides; inulin.  
 XX Penicillium purpurogenum.  
 OS JP07059574-A.  
 PN 07-MAR-1995.  
 XX 23-AUG-1993; 93JP-0229448.  
 XX 23-AUG-1993; 93JP-0229448.  
 XX (MITK ) MITSUI TOATSU CHEM INC.  
 PA WPI; 1995-135898/18.  
 DR N-PSDB; AAQ86692.  
 XX A new DNA encoding endo-inulinase - useful for the production of  
 PT inulo-oligosaccharides  
 XX Claim 1; Pages 4-5; 8pp; Japanese.  
 XX AAQ86692 encodes AAR72566 the Penicillium purpurogenum variety  
 CC rubrisclerotium endo-inulinase. The endo-inulinase can be used  
 CC to prepare an inulo-oligosaccharide from inulin.  
 XX SQ Sequence 490 AA;  
 Query Match 13.7%; Score 389.5; DB 16; Length 490;  
 Best Local Similarity 26.0%; Pred. No. 2.1e-25;  
 Matches 133; Conservative 87; Mismatches 225; Indels 67; Gaps 19;  
 QY 40 RAVYHMTTPSGWLCDPQRPVTHGAYQLYLLHSDQNN--GPGWDHASTTGDGVAFTTHGCT 97  
 Db 4 RPTFHCPAENWNEENGLIKIDSTWHLFYQADPTANVWNECHGATSSDLLHWHLPV 63  
 QY 98 VMLRPDPFVWSSAVVGTAANTAGGAGA---VVALATOPTDGVKRYQBYLYWSTDGDF 154  
 Db 64 AIPVENGIESFTGTSYDANNTSSLTSTNPPVLAFTGCTSS-NGTQDQRLAYSTDLCG 122  
 QY 155 TETALP-DPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVI--GELRYAAYF 211  
 Db 123 TWLKFSGNPII-----SAALEAPHDVVTGGLESKPKVFFHEPESKVMVLAHGGQDKLTWF 178  
 QY 212 TSNLNRDWTLRNF-----DYPNHALGGIECPDLFEITADDGTRH--WVL-----AASMD 259  
 Db 179 TSLDAKSWTMSDLLASQIEGPPS-SVTGNEVPDMFQLPI-QGTNETTWIIFTPAQGSP 236  
 QY 260 AVGIGLPWTYAYWTGWDGEQFHAD--DLTPQWLDGWWDYAAVTPSTDAPETRKLATA 317  
 Db 237 AGNGV-----VALTGSFDETFELANPVDSTLWLDYGRDFGAMSMENVPASDGRLLIAA 292  
 QY 318 WMNNKYAARDVPTDASDQYNGQNSIVRELRLARQPGGNYTLLSTPVAALTNYVTATTTL 377  
 Db 293 VMNSY---GSNPPTNT---WKGLSPRTLTLEKIGSKOY-FLOQPIAELSTVDNALASI 345  
 QY 378 PORTVDSAVLPWNGRAYEIEDIAMDATNVGISVGRSPDGRTHFNI--GKYGADLYVD 435  
 Db 346 ONQTIAPKOTLLSSIHGSSLDVRIAFSDVSGATLSLAVRKGSEQTVIRYSOSNSTLSVD 405  
 QY 436 RPSDLIAGYSLAPYSRAAAPIDPGARSVH-----LRILVDQSVEVFVNAGH 482  
 Db 406 RTASGDISY-----DPAAGGIHSAQLARDNTELVYLRVLVDTCSEVEFGQGE 453  
 QY 483 TVLSQOVHFAEGDGTGSLTYDGGPAHFTGIIV 514  
 Db 454 AVISDLIFFPSNSDGLSLEVIIGGTATLQSV 485  
 RESULT 12  
 AAY25346  
 ID AAY25346 standard; Protein; 750 AA.

XX AAY25346;  
 AC 03-SEP-1999 (first entry)  
 DT Bacillus sp. L7 endo-levanase protein.  
 XX Endo-levanase; L7; slime; control; industrial water; pulp; paper;  
 KW treatment; levan removal; industry.  
 XX Bacillus sp.  
 OS WO9931020-A1.  
 PN 24-JUN-1999.  
 XX 16-DEC-1998; 98WO-FI00987.  
 XX 12-JUN-1998; 98US-0089045.  
 PR 16-DEC-1997; 97US-0069801.  
 XX (CULT-) CULTOR CORP.  
 PA Miasnikov A;  
 PI WPI; 1999-405013/34.  
 DR N-PSDB; AAX78782.  
 XX Using an endo-levanase enzyme preparation  
 PT Disclosure; Page 18-21; 31pp; English.  
 XX This invention describes a novel method for the treatment of industrial  
 CC waters with an endo-levanase enzyme preparation which results in a  
 CC dramatically improved levan removal compared with current methods. The  
 CC method reduces the amount of slime present in industrial waters and  
 CC comprises contacting industrial water containing at least one slime  
 CC material with an endo-levanase enzyme preparation at a temperature and  
 CC pH condition sufficient to maintain activity of the enzyme. The  
 CC industrial waters treated are especially from pulp or paper processes.  
 XX SQ Sequence 750 AA;  
 Query Match 13.6%; Score 386; DB 20; Length 750;  
 Best Local Similarity 31.9%; Pred. No. 8e-25;  
 Matches 115; Conservative 44; Mismatches 144; Indels 58; Gaps 16;  
 QY 40 RAVYHMTTPSGWLCDPQRPVTHGAYQLYLLHSDQNNQNGPGWDHASTTGDVAFTHHGTVM 99  
 Db 398 ROYHYSPIRGSASDPNGLVYFEGEYHLP--HQD-----GGWAHVAVSRLIHWKRLPIAL 451  
 QY 100 PLRPDPFVWSSAVVGTANTAGF-----GAGAVVALATOPTDGVKRYQBYLYWSTDGDF 154  
 Db 452 PNDLGHVWSSGSAVADTTNASGLFGSSGKGLIAYTYSYNDPHNGNOKIGLAYSTDGR 511  
 QY 155 TTTALPD-PVIVNTDGRAATTPAEIENAEW-FRDPKIHWDTARGEWCVIGRLRYAAYT 212  
 Db 512 TWKYSSEHVPWITENPKTGEDP-----GGWDFRDPKVVREDEANNRWMVWVSGDHLRLET 566  
 QY 213 SPNLRDWTLRNFDYPNHALGGI-ECPLDLEITADDG-TRHNVLAASMDAYGIGLPMYTA 270  
 Db 567 STNLLNWTITDQFGYAYIRGGWECPCDLFQLPVSGSKRKKRWLVMISTGAN---PNTOG 622  
 QY 271 YWTGTWDSQFADHADDLPQW-----LQWGDWYAAVTPWPSIDAPETKRLATA 317  
 Db 623 S-----DAEYF-IGDLTPEGKFINDPACTVLKTDWKEYYASMSFS--DMDFGRIMLA 674  
 QY 318 WMNNKYAARDVPTDASDQYNGQNSIVRE-----LRLARQPGGNYTLLSTPVAALT 368  
 Db 675 WMTNNDYPP-SFPT---TCWKGQLSIPRQVSLKETEEGIRMHQTPTEELAQLRSPVLHIT 730  
 QY 369 N 369



Db 731 N 731

RESULT 13  
AAY24924

ID AAY24924 standard; Protein: 750 AA.  
XX  
AC AAY24924;  
XX  
DT 26-AUG-1999 (first entry)  
XX  
DE Bacillus L7 endo-levanase.  
XX  
KW Bacillus L7; endo-levanase; levan removal; sugar juice; sucrose;  
KW invertase; dextranase; filtration.  
XX  
OS Bacillus L7.  
XX  
PN W09931281-A1.  
XX  
PD 24-JUN-1999.  
XX  
PF 16-DEC-1998; 98WO-F100988.  
XX  
PR 16-DEC-1997; 97US-0069801.  
PR 16-DEC-1997; 97US-0069800.  
XX  
PA (CULT-) CULTOR CORP.  
XX  
PI Miasnikov A;  
XX  
DR WPI; 1999-395193/33.  
DR N-PSDB; AAX83630.  
XX  
PT Improving the filterability of sugar extraction juice with an  
PT endo-levanase enzyme  
XX  
PS Example; Page 19-22; 30pp; English.  
XX  
SS The present invention describes a method for improving the filterability  
CC of sugar extraction juice using an endo-levanase enzyme preparation  
CC devoid of sucrose (invertase) activity. The endo-levanase enzyme  
CC preparation is used (in combination with a dextranase) for treating the  
CC sugar extraction juice, specifically to remove levan. The selectivity  
CC of the endo-levanase enzyme solves the levan associated problems in the  
CC beet sugar industry. The present sequence represents Bacillus L7  
CC endo-levanase, for use in the method of the invention.  
XX  
SQ Sequence 750 AA;

Query Match 13.6%; Score 386; DB 20; Length 750;  
Best Local Similarity 31.9%; Pred. No. 8e-25;  
Matches 115; Conservative 44; Mismatches 144; Indels 58; Gaps 16;

Qy 40 RAVYHMTTPSGWGLCDPQRPVTHGAYQLYLHSDQNGPGGWDHASTTGDGVAFTHHGTVM 99  
Db 398 RPOYHSPIRGSASDPNGLVYFEGEYHLF--HQD---GGQWAAHVSRLDHWKRLPIAL 451  
Qy 100 PLRDPFPWSSGAVVGTANTAGF-----GAGAVVALATQPTDGVKRYOEQYLYWSTDGCF 154  
Db 452 PWNLDGHWGMSGAVADTTNAGSLFGSSGGKGLIAYTTSYNPDHNGNOKIGLAYSTDRGR 511  
Qy 155 TFTALPD-PVTVNTDGRAATTPAETENAEW-FRDPKIHWDTARGEWVCVIGRLRYAAPT 212  
Db 512 TWKYSEHPVVIENPGKTGEDP-----GGWDFRDPKVVDRDANRWVWVSGGDHRLUFT 566  
Qy 213 SPNLRDTRLRNFDPNHALGGI-ECPLDFEITADGG-TRHWVLAASMDAYGIGLPMPTYA 270  
Db 567 STNLLNWTLTDFQGYGAYIRGVWECPLDFOLPVEGSKRKWVLMISTGAN---PNTQG 622  
Qy 271 YWTGTWGEQPHADLLTPQW-----LDWGMWDTAAVTWPSIDAPETKRLAIA 317  
Db 623 S-----DAEYF-IGDLTPEGKFINDPAGTVLKTDMGCKEYASMSFS---DMPDGRRIMLA 674

Qy 318 WMNNWYKAARDVPTDASDQYNGQNSIVRE-----LRLARQPGGWYTLTSTPVAALT 368  
Db 675 WMTNWDYDF-SFPT---TGWKGLSIPROVSLKETEGERHMQHTPIEELAQLRSLVHIT 730

Qy 369 N 369  
Db 731 N 731

RESULT 14  
AAR56256  
ID AAR56256 standard; Protein: 556 AA.  
XX  
AC AAR56256;  
XX  
DT 17-FEB-1995 (first entry)  
XX  
DE Inulinase preprotein.  
XX  
KW Polymerase chain reaction; primer; amplify; PCR; extension assay;  
KW functionality; inulinase; promoter; regulatory region; promoter;  
KW upstream activating sequence; upstream repressing sequence; UAS; URS;  
KW terminator; secretory signal; secretion; yeast; transformation;  
KW protein; metabolite.  
XX  
OS Kluyveromyces marxianus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..16  
FT /note= "Pre-sequence"  
FT Binding-site 13..19  
FT /note= "Recognition site for signal peptidase"  
FT Cleavage-site 16..17  
FT /note= "Cleavage site for signal peptidase"  
FT Protein 17..556  
FT /note= "Mature Inulinase"  
FT Binding-site 22..23  
FT /note= "Processing site for KEX-2 like endoprotease"  
FT Cleavage-site 23..24  
FT /note= "Cleavage site for KEX-2 like endoprotease"

W09413821-A.  
23-JUN-1994.  
09-DEC-1993; 93WO-EP03547.  
11-DEC-1992; 92EP-0203932.  
29-JAN-1993; 93EP-0200240.  
(UNIL) QUEST INT BV.  
Chapman JW, Musters W, Rouwenhorst RJ, Toschka HY, Verbakel JMA;  
WPI; 1994-217894/26.  
N-PSDB; AAO66475.  
Yeast regulatory sequence from the inulinase gene - for protein,  
RNA or metabolite prodn.  
Disclosure; Fig 5; 68pp; English.  
This sequence represents the inulinase enzyme from K. marxianus. The  
DNA encoding this protein was obtained by using the two primer sequences  
given in AAO66460-61 to amplify a fragment corresponding to the  
inulinase 5' region and using the amplified fragment as a probe in the  
isolation of the full length sequence. In sequence analysis of the  
isolated gene, three dominant signals emerged coinciding with T-174,  
C-170 and C-167. These nucleotides are located approx. 100 nucleotides  
behind the TATA box. Association of the start of transcription with this  
region defines an uncommonly long 5' non-coding region. The MIG1 binding  
sequence perfectly matches the MIG1 binding site consensus sequence from



Search completed: May 6, 2003, 12:32:12  
Job time : 39.0673 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:30:07 ; Search time 29.0558 Seconds  
(without alignments)  
1723.790 Million cell updates/sec

Title: US-09-868-328B-1  
Perfect score: 2845  
Sequence: 1 MTPAISRRVAVLQAGAGALA.....TDGGAFTGIIVREIGQAI 521

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 2355  | 82.8        | 517    | 2 J5875  | levansucrase (EC 2 |
| 2          | 944   | 33.2        | 516    | 2 E70035 | levanase homolog y |
| 3          | 880.5 | 30.9        | 514    | 2 G97118 | levanase (imported |
| 4          | 562.5 | 19.8        | 512    | 2 A5206  | sucrose/fructan hy |
| 5          | 556   | 19.5        | 677    | 2 A27286 | levanase (EC 3.2.1 |
| 6          | 515   | 18.1        | 622    | 2 A36915 | fructanase - Bacte |
| 7          | 467.5 | 16.4        | 943    | 2 J4081  | sucrase/fructanase |
| 8          | 434   | 15.3        | 516    | 2 JEO301 | inulinase (EC 3.2. |
| 9          | 425   | 14.9        | 515    | 2 J47533 | inulinase (EC 3.2. |
| 10         | 406.5 | 14.3        | 1142   | 2 C97080 | levanase/invertase |
| 11         | 389.5 | 13.7        | 515    | 2 J65458 | inulinase (EC 3.2. |
| 12         | 382.5 | 13.4        | 1423   | 2 A49206 | exo-beta-D-fructos |
| 13         | 379   | 13.3        | 550    | 2 S6753  | beta-fructofuranos |
| 14         | 369.5 | 13.0        | 532    | 2 S27372 | beta-fructofuranos |
| 15         | 366.5 | 12.9        | 532    | 2 S27373 | beta-fructofuranos |
| 16         | 366   | 12.9        | 556    | 1 S31330 | inulinase (EC 3.2. |
| 17         | 365.5 | 12.8        | 555    | 1 S17502 | inulinase (EC 3.2. |
| 18         | 359.5 | 12.6        | 532    | 1 IFBY   | beta-fructofuranos |
| 19         | 355   | 12.5        | 477    | 2 C91034 | sucrose-6-phosphat |
| 20         | 354   | 12.4        | 477    | 2 S52162 | sucrose hydrolase  |
| 21         | 348   | 12.2        | 477    | 2 D85878 | sucrose hydrolase  |
| 22         | 348   | 12.2        | 581    | 2 J80102 | invertase - fissio |
| 23         | 330   | 11.6        | 476    | 1 C43717 | beta-fructofuranos |
| 24         | 320   | 11.2        | 533    | 2 S13528 | beta-fructofuranos |
| 25         | 318   | 11.2        | 490    | 2 T43149 | probable beta-fruc |
| 26         | 313.5 | 11.0        | 453    | 2 T39155 | glycosyl hydrolase |
| 27         | 312   | 11.0        | 590    | 2 T02096 | probable beta-fruc |
| 28         | 297.5 | 10.5        | 581    | 2 T12095 | beta-fructofuranos |
| 29         | 283.5 | 10.0        | 580    | 2 S49308 | beta-fructofuranos |

|    |       |     |     |          |                    |
|----|-------|-----|-----|----------|--------------------|
| 30 | 283   | 9.9 | 650 | 2 S23217 | beta-fructofuranos |
| 31 | 276   | 9.7 | 512 | 2 JU0460 | beta-fructofuranos |
| 32 | 274.5 | 9.6 | 636 | 1 S31157 | beta-fructofuranos |
| 33 | 274.5 | 9.6 | 636 | 1 S31155 | beta-fructofuranos |
| 34 | 272   | 9.6 | 479 | 2 S68598 | sucrose-6-phosphat |
| 35 | 262   | 9.2 | 584 | 2 T06163 | beta-fructofuranos |
| 36 | 261.5 | 9.2 | 582 | 2 S37047 | beta-fructofuranos |
| 37 | 260.5 | 9.2 | 640 | 2 T09534 | probable beta-fruc |
| 38 | 257   | 9.0 | 661 | 2 S37590 | beta-fructofuranos |
| 39 | 256   | 9.0 | 592 | 2 S61503 | beta-fructofuranos |
| 40 | 256   | 9.0 | 661 | 2 S37592 | beta-fructofuranos |
| 41 | 255   | 9.0 | 634 | 2 S31925 | beta-fructofuranos |
| 42 | 253.5 | 8.9 | 562 | 2 S7951  | beta-fructofuranos |
| 43 | 253   | 8.9 | 587 | 2 S36231 | beta-fructofuranos |
| 44 | 250.5 | 8.8 | 639 | 2 S71268 | beta-fructofuranos |
| 45 | 249   | 8.8 | 642 | 2 S55521 | beta-fructofuranos |

ALIGNMENTS

RESULT 1

JC5875

levansucrase (EC 2.4.1.10) precursor - Arthrobacter nicotinovorans

N:Alternate names: levan fructotransferase

C:Species: Arthrobacter nicotinovorans

C>Date: 09-Mar-1998 #sequence\_revision 09-Mar-1998 #text\_change 20-Jun-2000

C:Accession: JC5875; PC4508

R:Saito, K.; Yokota, A.; Tomita, F.

Biosci. Biotechnol. Biochem. 61, 2076-2079, 1997

A:Title: Molecular cloning of levan fructotransferase gene from Arthrobacter nicotino

A:Reference number: JC5875; MUID:98101974; PMID:9438987

A:Accession: JC5875

A:Molecule type: DNA

A:Residues: 1-517 <SA1>

A:Cross-references: DDBJ:AB001984; NID:g2897799; PIDN:BAA24915.1; PID:g2897800

A:Experimental source: strain GS-9

A:Accession: PC4508

A:Molecule type: protein

A:Residues: 34-45;137-166;185-238;309-319;424-445 <SA2>

C:Comment: This enzyme is involved in the production of levan which is a beta-2,6-li

C:Genetics:

A:Gene: lft

C:Superfamily: Penicillium purpogenum inulinase

C:Keywords: glycosyltransferase; hexosyltransferase

F:1-33/Domain: signal sequence #status predicted <SIG>

F:34-517/Product: levansucrase #status experimental <MAT>

F:52,232/Active site: Asp, Glu #status predicted

Query Match 82.8%; Score 2355; DB 2; Length 517;  
Best Local Similarity 81.0%; Pred. No. 2.7e-166;  
Matches 422; Conservative 41; Mismatches 54; Indels 4; Gaps 2;

QY 1 MTPAISRRVAVLQAGAGALALIFGGAVPPAARASAGSLRAVYHMTTPSGWCLDPPRPVT 60

Db 1 IYDIISRRALDQAGVAGALFMSNAIPVAAHAQA--SLRAIYHMTTPSGWCLDPPRPVH 58

QY 61 THGAYOLYLSHQDQNGPGGWDHASTDGVAFTHGTVMPDRDPFVWGSAGVVGTA 120

Db 59 TNGAYOLYLSHQDQNGPGGWDHASTDGVAFTHGTVMPDRDPFVWGSAGVVGTA 118

QY 121 GFGAGAVVALATQPTDGVKRYQEQYLYWSTDGGFTFTALPDPIVNTDGRAATPAEIN 180

Db 119 GFGAGAVIALATQPTDG--KFQEQYLYWSTDGGYSFTALPDPIVNTDGRATTAPEVN 176

QY 181 AEWFRDPKTHWDARGWCVCIGRLRYAAYFSPNLRDNLRFNFDYPNHALGGIECPDL 240

Db 177 AEWFRDPKTHWDATRNWCVCIGRLRYAAYFSPNLRDNLRFNFDYPNHALGGIECPDL 236

QY 241 FEITADGGTRHWVLAASMDAYGIGLPMYATGTGWDGQFHADDLTPQWLQWGDWYAA 300

Db 237 FEMTAGDGRHWVFGASMDAYSIGLPMYATGTGWDGQFHADDLTPQWLQWGDWYAA 296

QY 301 VTPSIDAPETKRLAIAMNNKYAARDVPTDASDGYNGONSIVRELRLARQPGGWYTL 360  
Db 297 VTPPAVEAPETKRLATAMNNKYAARNVPTDASDGYNGONSITRELRLERQSGGWYTL 356  
QY 361 STPVAALNTVYATTTLPDRDVGSAVLPNNGRAYEIELDIAWDTATNVGSGRSPDGT 420  
Db 357 STPVALSNATSTTLPDRDVGSAVLPNNGRAYEIELDIAWDTATNVGSGRSPDGT 416  
QY 421 RHTNIGKYGADLYVDRGSLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVFVNA 480  
Db 417 RHTNIGKYGDELVDYDRASSQSGVALAPYTRAAAPIDANARSVHLRLVLTQSVFVNA 476  
QY 481 GHTVLSQVHFAAGDGTISLYTGGPAHFTGIVVREIGQAI 521  
Db 477 GHTVLSQVHFAAGDGTISLYTGGPAHFTGIVVREIGQAI 517  
RESULT 2  
E70035  
levanase homolog yveB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: E70035  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Frenz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim  
wood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.  
Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber  
A:Authors: Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigu  
K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwol  
Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.J.; Schleic  
A:Authors: Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serron  
amakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol  
Wipat, A.; Yamamoto, H.; Yamakawa, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa,  
A:Authors: Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E70035  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Experimental source: strain 168  
C:Genetics:  
A:Molecule type: DNA  
A:Residues: 1-516 <KUN>  
A:Cross-references: GB:299121; GB:AL009126; NID:92635827; PIDN:CABL5451.1; PID:g2635959  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yveB  
C:Superfamily: Penicillium purporogenum inulinase  
Query Match 33.2%; Score 944; DB 2; Length 516;  
Best Local Similarity 41.8%; Pred. No. 4.9e-62;  
Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;  
QY 40 RAYVHTPPSGWGLCDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 97  
Db 44 RAYVHTPPSGWGLCDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 103  
QY 98 VMP--LRPDPVWSSGSAVGTANTAGAGAVVALATQPTDGVKRYQEOYLYWSTDGGFT 155  
Db 104 AIPKYNPDGDIWTSVVDKENTAGFGKALVAIVTQPSAKDKK-QEOYLYWSTDGGFT 162  
QY 156 F-----TALPDPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVVIGRLRYAA 209  
Db 163 FKPSYGNVMPNP---GTDD-----FRDPKIVWDDQDNKVMVMAEGSKIG 205  
QY 210 FYTSPNLRDWTLRNFDYPNHALGGIECPDLFEITADDGTRHVLAAASMDAYGIGLPMY 269  
Db 206 FYSDNLKMDHYTSGF-FPEQA-GMVECDPLVMRASDGTNKKVLGASANGKPKWGPNTY 263  
QY 270 AYTGTWDEQFHADDTLPQWLDGWGDWYAAVTPWSIDA--PETKRLAIAMNNKYAAR 327  
Db 264 AYTGTWDEQFHADDTLPQWLDGWGDWYAAVTPWSIDA--PETKRLAIAMNNKYAAR 322

QY 328 DVPTDASDGYNGONSIVRELRLARQPGGWYTLSTPVAALNTVYATTTLPDRDVGSAV 387  
Db 323 NTPT-MKNGFNGDTSVIRELRLKEQ-DGTYSLVSPQTEALEQLTSTDEIEDQDNGSKT 380  
QY 388 LPWNGRAYEIELDIAWDTATNVGSGRSPDGTFRHTNIGKY--CADLYVDRGSPSLAGYS 445  
Db 381 LSTIGDTQYQDTLDSSELNAGVRLRESDEQKRIHDVGFAGGYAYVNRATNQPKDS 440  
QY 446 LAPSRAAAPIDPGARSVHLRLVLTQSVFVNAAGTHTVLSQVHFAAGDGTISLYTGG 505  
Db 441 -NYVESKAPYDYNKRKVKHLKILVDKTTIEVFVGDGKTFVSNVFPKPKDGITLYSDGG 499  
QY 506 PAHFTGIVR 515  
Db 500 TASFKNITVK 509  
RESULT 3  
G97118  
levanase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: G97118  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
A:Authors: Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A95900; MUID:21359325; PMID:21359325  
A:Accession: G97118  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK79738.1; PID:g15024743; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1773  
C:Superfamily: Penicillium purporogenum inulinase  
Query Match 30.9%; Score 880.5; DB 2; Length 514;  
Best Local Similarity 40.7%; Pred. No. 2.4e-57;  
Matches 200; Conservative 76; Mismatches 179; Indels 37; Gaps 14;  
QY 40 RAYVHTPPSGWGLCDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 97  
Db 45 REYVHTFPNKKNDPQRPVTHGAYOLYXLS-DOINGPG-CWDHASTTGDVAFTHHGT 104  
QY 98 VMP--LRPDPVWSSGSAVGTANTAGAGAVVALATQPTDGVKRYQEOYLYWSTDGGFT 155  
Db 105 CIPKYNPDGDIWTSVVDKENTAGFGKALVAIVTQPSAKDKK-QEOYLYWSTDGGFT 163  
QY 156 FTALPD-PVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVVIGRLRYAAFTSP 214  
Db 164 FKPSYGNVMPNP---GTDD-----FRDPKIVWDDQDNKVMVMAEGSKIG 211  
QY 215 NLRDWTLRNFDYPNHALGGIECPDLFEITADDGTRHVLAAASMDAYGIGLPMYAYWTG 274  
Db 212 NLKNW--QHTGDFFTNIGVIECPDIFMOSDNGNTKWLGTSAANGKVSSEPTAYVWG 269  
QY 275 TWDEQFHADDTLPQWLDGWGDWYAAVTPWSIDA--KRLAIAMNNKYAARDVPTD 332  
Db 270 NYDKKFIADISTPKWLDYGFDMWIAVTESENQSNLKKRYALAMNNKY--ANWTPT- 327  
QY 333 ASDGYNGONSIVRELRLARQPGGWYTLSTPVAALNTVYATTTLPDRDVGSAVLPWNG 392  
Db 328 IQNRENGMSIVREITLSKQKNTYSLKPKIKRIENITTSIDQFQKISVKGLKHLKVGQ 387  
QY 393 RAYEIELDIAWDTATNVGSGRSPDGTFRHTNIGKY--CADLYVDRGSPSLAGYS LAP-- 448  
Db 388 KAYQDITDINWDNAKNGVRLRESLDRKRIHDVGFAGGYAYVNRATNQPKDS 441  
QY 449 ---YSRAAAPIDPGARSVHLRLVLTQSVFVNAAGTHTVLSQVHFAAGDGTISLYTGG 505  
Db 442 SKKIVESRAPFDLNNKVKHLRIFVDKVSVEVFIDDKITISNEVFPKPKDGITLYSIK 501



Db 258 MOYFVGDFGTHFKNENPPNKVLTMDYGRDYAAVSDIPSTDSRRRLWLGWMSWQY-A 316  
Qy 327 RQVPTDASDGYNGQNSIVRELRLARPGQWYTLTSTPVAALTNVYVATTTLPLDRTVDGSA 386  
Db 317 NDVPTSP---WRSATSPRELKLAFTGEG-VRVVQTPVKELETIRGTSKKWKNLTISPAS 372  
Qy 387 ---VLPWNGRAYEI--ELDIAMDTATNVGISVGRSPDGTGRTHTNIGKYGADLYVDRGPSDL 441  
Db 373 HNVLAGQSGDAYEINAEFKVSPGSAEFGFKVRTGENQFTKVGYYDRRNAKLFVDRSESGN 432  
Qy 442 AGYSLA-PYSRAAPIDPGARSVHLRLILYDTSQVEVFNAGHTVLSQOVHFAEGDTGISL 500  
Db 433 DTFNPAFTGKETAPLKEPVNGKVLRIYFVDRSSVEVEFGNDGKQVITDIIPLDRSSKGLEL 492  
Qy 501 YTDGGAHFTGIVVREI 517  
Db 493 YAANG-----GVKVKSL 504

RESULT 6  
A36915  
fructanase -- Bacteroides fragilis  
C:Species: Bacteroides fragilis  
C>Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 17-Mar-2000  
C:Accession: A36915  
R:Blatch, G.L.; Woods, D.R.  
J. Bacteriol. 175, 3058-3066, 1993  
A:Title: Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1  
A:Reference number: A36915; MUID:93259952; PMID:8491724  
A:Contents: BF-1  
A:Accession: A36915  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-622 <BLA>  
A:Cross-references: GB:M83774; NID:g143970; PIDN:AAA22924.1; PID:g143972  
A:Note: sequence extracted from NCBI backbone (NCBIN:131947, NCBI:P131949)  
C:Superfamily: beta-fructofuranosidase

Query Match 18.1%; Score 515; DB 2; Length 622;  
Best Local Similarity 30.2%; Pred. No. 3.1e-30;  
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

Qy 40 RAVYHMTTPSGWGLCDPQRPVTHGAYOLYLYHSDQNN--GPGGWDHASTTGDGVAFTHHGT 97  
Db 133 RPLVHTHTPLXGWMNDANGLVYKDEYHLFYQYNPYGSWMGMMHMGHSVSKDLVHWEH--- 189  
Qy 98 VMLRPDPF-----VMSGSAVVGCTANTAGGAGAVVALATQPTDGYRKYQEQYLYHSTD 151  
Db 190 ---LEPALARDTLGHIFSGSSVVDANTAGYAGAIYAFYTSASD---KNGOIOCMAYSTD 244  
Qy 152 GGFFTTAL-PDPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEVWCVIGRLRYAAF 210  
Db 245 NGRTFTYKRNPLVTPDQ-----LKDPRDKPVFYAPDQKWVWVNSADKEMRF 293  
Qy 211 YTSNLRDWTLLRRNF-----YPNHALGIECPDLFEITADGTRH--WVLAASMDAYGI 263  
Db 294 YSSENLEWYMSGWEGYGVQPSQ-----PECPDMVELPVDGNPDHKKWALIVNVN--- 345  
Qy 264 GLPMTY-----AYWTGTWGEQFHADDL--TPQWLDWGDWDYAAVTPSIDAPETKRLA 315  
Db 346 --PGCYFGSATQYFIDFGDGEKFCVCKNPETVKWLDWGDWHDYATVCFSTNG---DRTIA 400  
Qy 316 IAWNNKRYAARDVPTDASDGYNGQNSIVRELRLARPGQWYTLTSTPVAALTNVYVATT 375  
Db 401 VPWMSNMQY-ANIVPTPQ---FRSANAALPRELSLYTQGDYI-MAAAPVEETKSLRKESR 455  
Qy 376 TLDPDRTVDGS-----AVLPWNGRAYEIELDIAMDTATNVGISVGRSPDGTGRTHTNIGKYGAD 431  
Db 456 EIPAFEGDAYHVDLSLSDNKGAYEIELELAAGSAEIMGLKLFNEKGENDVIVISLPEKK 515  
Qy 432 LYVDRGPSDLA--GYSLAPYSRAA-----APIDPCARSVHLRLV 469

Db 516 LVMDRTKSGIVDFGKDSAPHAEIAHRRKQNSINVYDDFALGTWAPVQK-AGNYKLDIEV 574  
Qy 470 DTQSVSEVFNAGHTVLSQOVHFAEGDTGISLYTDCG 505  
Db 575 DKCSVEIFLNGGKIATNLIFPTTPYQMSFYSRGG 610

RESULT 7  
JC4081  
sucrase/fructanase precursor - Actinomyces naeslundii (strain T14V)  
C:Species: Actinomyces naeslundii  
C>Date: 30-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 15-Oct-1999  
C:Accession: JC4081  
R:Norman, J.M.; Bunney, K.L.; Giffard, P.M.  
Gene 152, 93-98, 1995  
A:Title: Characterization of levJ, a sucrase/fructanase-encoding gene from Actinomyces  
A:Reference number: JC4081; MUID:95129923; PMID:7828936  
A:Accession: JC4081  
A:Molecule type: DNA  
A:Residues: 1-943 <NOR>  
A:Cross-references: GB:U12274; NID:g515689; PIDN:AAA67876.1; PID:g515690  
A:Note: The authors translated the codon ATG for residue 467 as Asn, TAC for residue  
C:Genetics:  
A:Gene: levJ  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-943/Product: sucrase/fructanase #status predicted <MAT>

Query Match 16.4%; Score 467.5; DB 2; Length 943;  
Best Local Similarity 21.5%; Pred. No. 1.8e-26;  
Matches 187; Conservative 83; Mismatches 217; Indels 381; Gaps 27;

Qy 4 AISRAVLOGAGAGALALIFGG-----AVPPAARASAPGS-LRAYVHMTTPSGWGLCDPQRP 58  
Db 19 AASTLASMPAAASGQARPVSAAAPNAPAPATAPKADQGTGERWRPQSHVTPKQNMNDPNGL 78  
Qy 59 VTHGATQLYLYHSDQNN--GPGGWDHASTTGDGVAFTHHGTVMPLRDPDFPWSSSAVVG 116  
Db 79 VYDGEYHMFYQYNPYGSWMGMMHMGHSVSKDLVHWEHVAIPHTSOYGVGFSASVIDT 138  
Qy 117 ANTAGFCA---GAVVALATQPTDGYRKYQEQYLYHSTDGDTGFTTAL--PDPVIVNTDGRA 171  
Db 139 KNTSGLSGSPDNPAWVAVWTRADYVGN--QSQSLAYSTDGKGTWNLXNGDPVL----- 189  
Qy 172 ATTPAEIENAEWFRDPKIHWDTARGEVWCVIGRL--RYAAPYTSNLRDWTLLRRNFYDN 229  
Db 190 -----DIGSNE-FRDPKVFWDQASGRWTVVSHATEHRVSEYSPDLIHMTQEVSFGGEG 243  
Qy 230 HALGGIECPDLFEITADGTRH--WVLA----- 255  
Db 244 ITSAVWACPDFFPLPVDGSSQEQVKWLVTVVADSQAQYFVGSWDCGTTFTPDEIPHYSGE 303  
Qy 256 -----ASMDAYG----- 262  
Db 304 TTLADFENGYAGWKADGAAGFGSGPATGDLPGHQKAYVDSFGSGDADTGLTSDFTVSS 363  
Qy 263 ----- 262  
Db 364 SYINLRTAGGKHPYNQATGONGGRLLAGFDGSGWEGTVEGSAFAATPPQATPAQOPL 423  
Qy 263 -----ICLPMTY-----AY----- 271  
Db 424 VNHSAGLLNTYLDAAATGOGSDAPTGTATGPTFTTIDSAIYNLNLGNGNNRPRREGADGGS 483  
Qy 272 -----W-----TGTW----- 276  
Db 484 RVSVELIVDGVKVSATGRNLEELNQSDVSLKCKSAQIVVTDATGCGWHILLDEV 543  
Qy 277 ----- 276  
Db 544 RASDKASPIADNTSVNLVVDGVKVASATGNNSCGLEWTSNMVAAAYKGRKARLVIEDRNG 603  
Qy 277 -----DGEQFHADDLTPQWLDWGDWYAAVTPSIDAPETKRLIAWMN 320



Db 604 NAEDGHLMVDDILOLSTKAFSGADVPR-LDYGKDYAAVTDWNV--PNGKRYQVGMWS 660  
QY 321 NMYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLTTPVAALTNYVTA-----T 374  
Db 661 NWAY-VRDLPITT---WRTAMSTVREMGLTR-VNGKRLRTAQVTALESRLTQDELIRKD 715  
QY 375 TTLP-DRYVDGSAVLPNWNGRAYEIELDAMDATNVNGISVGRSPDGTHTNTG--KYGAD 431  
Db 716 TDIPVGETSLGKAA---QCTSLDLSVDLSPSASSFAGLVK--LDNGEQVTLIGYDSQAQK 770  
QY 432 LVYDRGSDLAGYSLAPYSRAAAPDPGAR-SVHLRLVDTOSVEFVFNAGHTVLSQQVH 490  
Db 771 LVYDRGSGVTDSPKFFARSTAPLSPDSKQGVHLRLIIVDAHSVEFAADGTPVITQTVY 830  
QY 491 FAEGDTGISLYDGGPAHFTGIWVREIG 518  
Db 831 PRQDATGVSLYAEAGCTAHLGSLSLWHLG 858  
  
RESULT 8  
JE0301  
inulinase (EC 3.2.1.7) - Aspergillus niger  
C:Species: Aspergillus niger  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 07-May-1999  
C:Accession: JE0301  
R:Ohta, K.; Akimoto, H.; Matsuda, S.; Toshimitsu, D.; Nakamura, T.  
Biosci. Biotechnol. Biochem. 62, 1731-1738, 1998  
A:Title: Molecular cloning and sequence analysis of two endoinulinase genes from Aspergillus niger  
A:Reference number: JE0301; MUID:99022189; PMID:9805373  
A:Accession: JE0301  
A:Molecule type: mRNA  
A:Residues: 1-516 <OHT>  
A:Cross-references: DBJ:AB012771  
C:Superfamily: Penicillium purporogenun inulinase  
C:Keywords: glycosidase; hydrolase  
  
Query Match 15.3%; Score 434; DB 2; Length 516;  
Best Local Similarity 27.1%; Pred. No. 2.3e-24;  
Matches 144; Conservative 92; Mismatches 22; Indels 70; Gaps 19;  
  
QY 24 CGAVPPAARASAPGSLRAYHMTTPSGWCLCDPQRPVTHGAYQLYLLHSDQNNQNGPGG--W 81  
Db 16 GLMLPSQAQSN---DYRPSYHFTPDQYWMNEPNGLIKGTWHLFPQHNPNTANVWNCIW 72  
QY 82 DHAISTDGVAFTHGCTVPLRDPFVWGSAAVVGTAANTAGFGAGA---VVALATOPTDGV 138  
Db 73 GHATSTDLHWAHAKFTAADENGVEAFTGTAYIDPNTSGLSDANPPYLAFTGYTTS- 131  
QY 139 RYQEOYLWSPDGGFTTALPDVPVINTDGRAATTPAEIENAEWFRDPKIHWDTARGEW 198  
Db 132 SOTQORLAFSDNGATWTKFOGNIISTSQEA---PHDITGLESRDPKVFFHRSQGNW 188  
QY 199 VCVI--GRLRYAAFTSPNLRDWTLRNFDYPNHALGGI-----ECPDLEIITADGCT 249  
Db 189 INVLAHGODKLSFWTSADTINWTSQ---DLKSTSLNGSSDITGWEPDMEFLPV-EGT 245  
QY 250 RH--WVL---AASMDAYGIGLPMYAYWGTGWDGEQHAD--DLTPQWLDGWDWYAAV 301  
Db 246 BETTWVMMTPAEGSPAGNGVLAI-----TGSFDEKSTADPVDASTWLDNGRDFDGL 301  
QY 302 TWPSIDAPETKRLATAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLIS 361  
Db 302 SWNVNPSADGRRIIAAVNSY---GSPNPTTT---WKGMLSPRTLSL-KKVGTOQHFFVQ 354  
QY 362 TPVAALTNYVATTLPDRTVDGSVAVLPWNGRAYEIELDIW--DTATNVGISVGRSPDG 419  
Db 355 QPITELDPTISLQTLANOTITPGQTLSSIRGTALDVRVAFYPDAGSVLSLAVRKGASE 414  
QY 420 TRHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVH-----LR 466  
Db 415 QTVINYTQSNATLSVDRFESGDISY-----DPAAGGVHTAKLEEDGCTGLVSIR 462  
QY 467 ILVDTQSVEFVFNAGHTVLSQQVHFAEGDGTGISLYTDGGPAHFTGIWVREI 517

Db 463 VLVDTCSEVFGQGEAVISDLIFPSSDSDGLALEVTGCGNAVLQSVDDVRSV 513

#### RESULT 9

JE07533

inulinase (EC 3.2.1.7) - Penicillium sp. (strain TN-88)

N:Alternate names: 2,1-beta-D-fructan fructanohydrolase

C:Species: Penicillium sp. (strain TN-88)

C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001

C:Accession: JC7533

R:Akimoto, H.; Kiyota, N.; Kushima, T.; Nakamura, T.; Ohta, K.

Biosci. Biotechnol. Biochem. 64, 2328-2335, 2000

A:Title: Molecular cloning and sequence analysis of an endoinulinase gene from Penicillium sp.

A:Reference number: JC7533; MUID:21036892; PMID:11193399

A:Accession: JC7533

A:Molecule type: mRNA

A:Residues: 1-515 <AKI>

A:Cross-references: DBJ:AB041337

A:Experimental source: strain TN-88, wild type

C:Comment: This enzyme, an acidic glycoprotein and an endo-acting inulinase as well, 1

as well as the main products.

C:Genetics:

A:Gene: inuC

C:Superfamily: Penicillium purporogenun inulinase

C:Keywords: glycoprotein; glycosidase; hydrolase

Query Match 14.9%; Score 425; DB 2; Length 515;

Best Local Similarity 27.6%; Pred. No. 1.1e-23;

Matches 142; Conservative 92; Mismatches 232; Indels 48; Gaps 21;

QY 31 ARASAPGSLRAYHMTTPSGWCLCDPQRPVTHGAYQLYLLHSDQNN--GPGGWDHASTD 88

Db 21 ARAVA--DYRPAFHCPAENWNEPNGLLIQINSTWHLFYQADPAANWGCWGHATSSD 79

QY 89 GVAFTHHGTVMPLRDPFVWGSAAVVGTAANTAGFGAGA---VVALATQDGTGVKRYQEQY 145

Db 80 LLHWDLHPVAIPVENGIESFTGTSYDSNNTSGLTSTNPPYLAFTGYTES-NKTQDQR 138

QY 146 LYKSTDGGTFTALP--DPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVI-- 202

Db 139 LAYSTDLGQVTKVFAGNPII---GAAQEPQDISGLESRDPKVFFHAPSCKWVMVLAH 194

QY 203 GLRLYAAFTSPNLRDWTLRNFD-----YPNHALGIECPDLFEITADGTRH--WVL 254

Db 195 GQDKLITFTSLDAKNWTVSDLSSSQIBGFPS-SITGWEVPMFOLPI-QGIKKTTWL 252

QY 255 ----AASMDAYGIGLPMYAYWGTGWDGEQHADLTPO--WLDGWDWYAAVTPSIDA 308

Db 253 IFTPAQGSPPGGNV---VALTGSFDGETFVADPVPSTLWLDYGRDFGALSWEVPA 308

QY 309 PETKRLATAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGWYTLTTPVAALT 368

Db 309 SDGRILIAAVMSY---GSPNPTTT---WKGMLSPRTLAL-KQISKEYFLOQPAEUS 361

QY 369 NYVTATTTLPDRTVDGSVAVL--PWNGRAYEIELDIWDTATNVGISVGRSPDGTHTNIC 426

Db 362 TIDGSLTSTQNTIIPNQTLLSSIHGTSIDIRMAFVIDSGATLSLAVRKG--GSEQTVIR 419

QY 427 KY--GADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEFVFNAGHT 483

Db 420 YFQSNSTLSVDRITASDISYDPAAGGVHTAQLAQDNTELHVHIALIDTCSVEFVGEGBA 479

QY 484 VLSQQVHFAEGDGTGISLYTDGGPAHFTGIWVREI 517

Db 480 VISDLIFPSSDSDGLSLEVSGGTAMLRSVNVSSV 513

#### RESULT 10

C97080

levanase/invertase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

Db 800 LAITPLKSAV 809

RESULT 11

JC5458

inulinase (EC 3.2.1.7) precursor - Penicillium purporogenum  
N:Alternate names: 2,1-beta-D-fructan fructanohydrolase  
C:Species: Penicillium purporogenum  
C:Date: 17-Jun-1997 #sequence\_revision 17-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: JC5458; PC4326  
R:Onodera, S.; Murakami, T.; Ito, H.; Mori, H.; Matsui, H.; Honma, M.; Chiba, S.; Shi  
Biosci. Biotechnol. Biochem. 60, 1780-1785, 1996  
A:Title: Molecular cloning and nucleotide sequences of cDNA and gene encoding endo-in-  
F:26-515/Product: endo-inulinase #status predicted <SIG>  
A:Reference number: JC5458; MUID:97141595; PMID:8987853  
A:Accession: JC5458  
A:Molecule type: DNA  
A:Residues: 1-515 <ON01>  
A:Cross-references: DDBJ:D84360; NID:g1805497; PIDN:BAI2321.1; PID:g1805498  
A:Accession: PC4326  
A:Molecule type: protein  
A:Residues: 26-70;152-239;333-345;350-379 <ON02>  
C:Comment: This enzyme hydrolyzes inulin and fructo-oligo-saccharides, but not sucros  
C:Superfamily: Penicillium purporogenum inulinase  
F:Keywords: glycoprotein; glycosidase; hydrolase  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-515/Product: endo-inulinase #status predicted <MAT>  
F:108,109,132,246,424,489/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 13.7%; Score 389.5; DB 2; Length 515;  
Best Local Similarity 26.0%; Pred. No. 4.5e-21;  
Matches 133; Conservative 87; Mismatches 225; Indels 67; Gaps 19;

Qy 40 RAVYHMPSPGSLCDPQRPVTHGAYOLYLHSDNN--GPGGWDHASTTGDGVAFTHHGT 97  
Db 29 RPTFHCPCAEWNMEPNGLIKIDSTWILFQADPTANVNGECWGCHATSDDLHWDHLPV 88  
Qy 98 VMPLRPDPFVMSGAVVGTANTAGFGAGA---VVALATQPTDGVRYQEOYLXWSTDGGF 154  
Db 89 AIPVENGIESFTGTSYVDANNTSSLGSTNPPYLAFPTGYTSS-NGTQDQRLAYSTD LGT 147  
Qy 155 TFTAALP--DPVIVNTDGRAATTPEIAENAFRDPKIHWDTARGEWCVI--GRLRYAAFY 211  
Db 148 TWLFKSGNPIT---SAALEAPHDVTGGLESRDPKVFHFEPGSKVMVLAHGQDGLTFW 203  
Qy 212 TSPNLRDWTLLRRNF-----DYPNHALGGTECPDLPEITADDGTRH--WVL----AASMD 259  
Db 204 TSLDAKSWTWMSDLLASQIEGFPS-SVTGNEVPDMFOLPI-QGTNETTWIITPAQGSP 261  
Qy 260 AYGIGLEPMTYAYWTGWDGQFHAD--DLTPQWLDMGWDWYAAVTPSIDAPETKRLAIA 317  
Db 262 AGNGV---VALTGSFDETFLLANPVDSSSTLWLDYGRDFDGAMSWENVPASDGRLLIAA 317  
Qy 318 WMNNWKYAARDVPTDASDGYNGQNSIVRELRARQPCGWYLLSTLTPVAALTNNVYATTTL 377  
Db 318 VMNSY---GSNPPTNT---WKGMLSPFRTLTLEKIGSKQY-FLQOPIAELSTVDNALASI 370  
Qy 378 PDRIVDGSAVLPWNGRAYEILEDTANDTATNVGISVGRSPDGTRHTNI--GKYGADLYVD 435  
Db 371 ONQTIAPKQTLILLSIHGSSLDVRIATSVDSGATLSLAVRKGSGEQVIRYSQSNSTLSVD 430  
Qy 436 RGPSDLAGYSIAPYSRAAAAPIDPCARSVH-----LRLIVDTQSVFVFNAGH 482  
Db 431 RTASGDLSY-----DPAAGGIHSAQLARDNTELVLVRLVLDVTSVEVFGGGE 478  
Qy 483 TVLSQQVHFAGDGTGSLYTDGGPAHFTGIW 514  
Db 479 AVISDLIFPSSNDGLSLEVGIGTATLQSV 510

RESULT 12

A49206

exo-beta-D-fructosidase - Streptococcus mutans  
C:Species: Streptococcus mutans

A/Accession: A49206  
A/Status: preliminary  
A/Molecule type: nucleic acid  
A/Residues: 1-1423 <BUR>  
A/Cross-references: GB:L03358; NID:g4733888; PID:AAA26889.1; PID:g153634  
A/Note: sequence extracted from NCBI backbone (NCBIN:116439, NCBIP:116440)

140 KYQEQLYLNWSTDGGFTFTALPDPIVYVNTDGRAATTPAEIENAEWFRDPKIHWDTARGET

QY 253 VLAASMDAYGIGLPM---TYAYWTGTWGDGEQFHADDLTPQWLDWGWDWYAAVTVPSID

```

QY 366 ALTNYTATTTL-----PDRTVDGSAVLPNWNGRAYE-----IELDIAW
      | | | | | : | | : : : : : | | | | | : | | :
DH 356 ---VDRTPRPI NERTI KTI ETI NEVEVNDI K FOK SSEVATNTENTENATCVFFFI KE

```

**QY**   **439**   SDLAGYSIAPYRAAIPIDPGARSVH-LRILVDTSVEVFVNAGHTVLSSQQVHFREG-

| : :: | | | | | : || :|| :|| :| :

**Dδ**   **466**   SPVFTERISTYVEKIDTTDQG--NVYTLYGVVDNRNILELFDNDGSIAMTNTFFFFREGK

RESULT 14  
S27372

R: Hohmann, S.  
submitted to the EMBL Data Library, November 1988  
A: Reference number: S27372

Mol. Gen. Genet. 211, 446-454, 1988  
A>Title: Structural analysis of the 5' regions of yeast SUC genes re  
A:Reference number: S25439; MUID:88216256; PMID:2835632.

A: Molecule type: DNA  
A: Residues: 1-74 <HO2>

```
Query Match      13.3%; Score 379; DB 2; Length 550;
Best Local Similarity 24.6%; Pred. No. 3e-20;
Matches 139; Conservative 87; Mismatches 212; Indels 128;

Qy 39 LRAVYHMTPSGWLCDPQPVTTHGAY-----QLYYLHSQDNGN-----PGGWDRHA
```

140 KYQEQLYLNWSTDGGFTFTALPDPIVYVNTDGRAATTPAEIENAEWFRDPKIHWDTARGET

QY 253 VLAASMDAYGIGLPM---TYAYWTGTWGDGEQFHADDLTPQWLDWGWDWYAAVTVPSID

```

QY 366 ALTNYTATTTL-----PDRTVDGSAVLPNWNGRAYE-----IELDIAW
      | | | | | : | | : : : : : | | | | | : | | :
DH 356 ---VDRTPRPI NERTI KTI ETI NEVEVNDI K FOK SSEVATNTENTENATCVFFFI KE

```

**QY**   **439**   SDLAGYSIAPYSRAAPIDPGARSVH-LRILVDTSVEVFVNAGHTVLSSQQVHFREG-

               | : :: |   | | : || :|| :| :| :| :

**D6**   **466**   SPVFTERISTYVEKIDTTDQG--NVYTLYGVVDNRNILELFENDGSGSIAMTNTFFFFREGK

RESULT 14  
S27372

R: Hohmann, S.  
submitted to the EMBL Data Library, November 1988  
A: Reference number: S27372

Mol. Gen. Genet. 211, 446-454, 1988  
A>Title: Structural analysis of the 5' regions of yeast SUC genes re  
A:Reference number: S25439; MUID:88216256; PMID:2835632.

A: Molecule type: DNA  
A: Residues: 1-74 <HO2>

A;Cross-references: EMBL:X07570

C;Genetics:

A;Gene: SUC1

A;Map position: 7R

C;Superfamily: beta-fructofuranosidase

C;Keywords: glycosidase; hydrolase

F;1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F;20-532/Product: beta-fructofuranosidase 1 #status predicted &lt;MAT&gt;

Query Match 13.0% Score 369.5; DB 2; Length 532;  
Best Local Similarity 25.1%; Pred. No. 1.4e-19;  
Matches 140; Conservative 84; Mismatches 221; Indels 113; Gaps 25;

QY 19 LALIFGAVPPAARASAPGSLRAVYHMTTPSGWLCDPQR--PVTHGAYOLYLHSDQNN 76

DQ 7 LFLLAGFAAKISASMTNETSDRPLVHFTPNKGMNDPGLWDYDAKEGKWHLYFYQNPNDT 66

QY 77 G---PGGWDHASTTDCGVAFTHHG---TVMPRLRDPFVWSSGSAVGTANTAGCAGAV-- 127

DQ 67 VWGLPLFWGHATSD---LTHWQDEPVIAIAKPKKDSGAYSGSMVIDYNNYSGFNDTIDP 123

QY 128 ---VALATOPTDCGVKRYQOYLWSTDCGFTFTAL--PDPVIVNTDGRAATTAEIENAE 182

DQ 124 RQRCVAIWTYNT--ESEEQYISYSLDGGYTFTEYQKNPVL-----ANST 167

QY 183 WFRDPKIHMTDARGWCVIGRLR--YAAFYTSPNLRDWTLRNFDPYFNHALGG--IECP 238

DQ 168 QFRDPKVFWEPSKKIWTAAKSDYKIEIYSSDDLKSWKLESAP--ANEGFLGYQYEC 225

QY 239 DLEFITA--DDGTRHWVLAASMDAYGICLPMTYA---YWTGTWGDGEFHADDTLPQWLDW 293

DQ 226 GLIEVSEQDPKSKHWMFISINP---GAPAGGSFNQYFVSGNHFHFAFDNQSRVYDF 282

QY 294 GWDYAAVTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGNSIVRELRLAROP 353

DQ 283 GKDYALQTFENTDPTGSGALGIAMNSWEYSA-FVP---SNPWRSSMSLVRFPSLNT- 337

QY 354 GGWYTLSTPVAALTNNVYATTTLPDRTVDGSAVLPN-----GRAYEIELDIAD 403

DQ 338 ---YQANPETELINLKA-----EPILNITSSAGPWSREATNTTLTKANSYVNLNST 386

QY 404 DT-----ATNVGISVGRS-----PDGTRHNTGKYADLYVDRGPSDLA 442

DQ 387 GTLEFELVAVNTQTTSKSVFADLSLWFKGLEDPPEYLRMGFEVSASSFFLDGRNSKVK 446

QY 443 GYSLAPY--SRAAAPIDPGARSVHLRI-----LVDTSQSVFVNAGHTVLSQQVHFAEGD 495

DQ 447 FVENPFTNRMSSVNNQPFKSENDLSYKYVYGLLDQNLILELYFNDGDVVSNTYFMTTGN 506

QY 496 -----TGIS--LYTD 503

DQ 507 ALGSVNNMTTGVNDLFLYID 524

## RESULT 15

S27373

beta-fructofuranosidase (EC 3.2.1.26) 4 precursor - yeast (Saccharomyces cerevisiae)

N;Alternate names: invertase; saccharase

C;Species: Saccharomyces cerevisiae

C;Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 22-Jun-1999

C;Accession: S27373; S25442

R;Hohmann, S.

submitted to the EMBL Data Library, November 1988

A;Reference number: S27372

A;Accession: S27373

A;Molecule type: DNA

A;Residues: 1-532 &lt;HOH&gt;

A;Cross-references: EMBL:X07572; NID:g4572; PIDN:CAA30459.1; PID:g4573

R;Hohmann, S.; Gozalbo, D.

Mol. Gen. Genet. 211, 446-454, 1988

A;Title: Structural analysis of the 5' regions of yeast SUC genes revealed analogous pal

A;Reference number: S25439; MUID:88216256; PMID:2835632

A;Accession: S25442

A;Molecule type: DNA

A;Residues: 1-74 &lt;HOR2&gt;

A;Cross-references: EMBL:X07572

A;Note: the authors translated the codon ATT for residue 7 as Asn

C;Genetics:

A;Gene: SUC4

C;Superfamily: beta-fructofuranosidase

C;Keywords: glycosidase; hydrolase

F;1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F;20-532/Product: beta-fructofuranosidase 4 #status predicted &lt;MAT&gt;

Query Match 12.9% Score 366.5; DB 2; Length 532;  
Best Local Similarity 26.2%; Pred. No. 2.4e-19;  
Matches 143; Conservative 80; Mismatches 231; Indels 91; Gaps 24;

QY 21 LIFGGAVPAAARASAPGSLRAVYHMTTPSGWLCDPQR--PVTHGAYOLYLHSDQNNG- 77

DQ 9 LLAGFAAKISALMTNETSDRPLVHFTPNKGMNDPGLWYDAKEGKWHLYFYQNPNDTV 68

QY 78 --PGWDHASTTDCGVAFTHHG---TVMPRLRDPFVWSSGSAVGTANTAGCAGAV---- 127

DQ 69 GLPLFWGHATSD---LTHWQDEPVIAIAKPKKDSGAYSGSMVIDHNTSEFFNDTVDPR 125

QY 128 --VALATOPTDCGVKRYQOYLWSTDCGFTFTAL--PDPVIVNTDGRAATTAEIENAE 184

DQ 126 RCVAIWTYNT--ESEEQYISYSLDGGYTFTEYQKNPVL-----ANSTQF 169

QY 185 RDPKIHMTDARGWCVIGRLR--YAAFYTSPNLRDWTLRNFDPYFNHALGG--IECPDL 240

DQ 170 RDPKVFWEPSKKIWTAAKSDYKIEIYSSDDLKSWKLESAP--ANEGFLGYQYEC 227

QY 241 FEITA--DDGTRHWVLAASMDAYGICLPMTYA---YWTGTWGDGEFHADDTLPQWLDW 295

DQ 228 IEVPTEDQPSKSHWMFISINP---GAPAGGSFNQYFVSGNHFHFAFDNQSRVYDF 284

QY 296 DWYAAVTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGNSIVRELRL-----A 350

DQ 285 DYVALQTFENTDPTGSGALGIAMNSWEYSA-FVPTNP---WRSSMSLVRFPSLNTYQA 340

QY 351 RQPGGWYTLSTPVAALTNNVYATTTLPDRTVDGSAVLPN-----GRAYEIELDIAD 405

DQ 341 NPETELINLKAEPILNITSSAGPWSREATNTTLTKANSYVNLNST 386

QY 406 ATNVGISVGRS-----PDGTRHNTGKYADLYVDRGPSDLA 442

DQ 400 TQSVKSVFSDLSLWFKGLEDPPEYLRMGFEVSASSFFLDGRNSKVK 446

QY 454 APIDPGARSVHLRI-----LVDTSQSVFVNAGHTVLSQQVHFAEGD- 499

DQ 460 VNNOPFKSENDLSYKYVYGLLDQNLILELYFNDGDVVSNTYFMTTGN 519

QY 500 -LYTD 503

DQ 520 LFYID 524

Search completed: May 6, 2003, 12:35:46

Job time: 33.0558 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:52 ; Search time 10.5202 Seconds

(without alignments)

2054.066 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845

Sequence: 1 MTPAIRRRRAVLQAGAGALA.....TDGGPAHFTGIVREIGQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Query | Score | Match | Length | ID         | Description        |
|------------|-------|-------|-------|--------|------------|--------------------|
| 1          | 556   | 19.5  | 677   | 1      | SACC_BACSU | P05656 bacillus su |
| 2          | 382.5 | 13.4  | 1423  | 1      | FRUA_STRMU | Q03174 streptococ  |
| 3          | 379   | 13.3  | 550   | 1      | INV1_HANAN | P40912 hansenua a  |
| 4          | 369.5 | 13.0  | 532   | 1      | INV1_YEAST | P10594 saccharomyc |
| 5          | 366.5 | 12.9  | 532   | 1      | INV4_YEAST | P10596 saccharomyc |
| 6          | 365.5 | 12.8  | 555   | 1      | INV1_KLUMA | P28999 kluyveromyc |
| 7          | 359.5 | 12.6  | 532   | 1      | INV2_YEAST | P00724 saccharomyc |
| 8          | 354   | 12.4  | 477   | 1      | CSCA_ECOLI | P40714 escherichia |
| 9          | 348   | 12.2  | 581   | 1      | INV1_SCHPO | O59852 schizosacch |
| 10         | 330   | 11.6  | 476   | 1      | RAFD_ECOLI | P16553 escherichia |
| 11         | 320   | 11.2  | 533   | 1      | INV1_DEBOC | P24133 debaryomyce |
| 12         | 312   | 11.0  | 590   | 1      | INV4_MAIZE | P49174 zea mays (m |
| 13         | 276   | 9.7   | 512   | 1      | INV4_ZYMMO | P35636 zymomonas m |
| 14         | 274.5 | 9.6   | 636   | 1      | INV4_LYCES | P29000 lycopersico |
| 15         | 260.5 | 9.2   | 640   | 1      | INV1_CAPAN | P93761 capsicum an |
| 16         | 249   | 8.8   | 642   | 1      | INV4_VICFA | Q43857 vicia faba  |
| 17         | 248   | 8.7   | 651   | 1      | INV4_PHAVU | O24509 phaseolus v |
| 18         | 246.5 | 8.7   | 465   | 1      | SCRB_KLEPN | P27217 klebsiella  |
| 19         | 246.5 | 8.7   | 555   | 1      | INV1_PEA   | Q43089 pisum sativ |
| 20         | 246   | 8.6   | 661   | 1      | INV2_DAUCA | P80065 daucus caro |
| 21         | 243   | 8.5   | 670   | 1      | INV1_MAIZE | P49175 zea mays (m |
| 22         | 242   | 8.5   | 592   | 1      | INV1_DAUCA | P26792 daucus caro |
| 23         | 242   | 8.5   | 649   | 1      | INV4_PHAUA | P29001 phaseolus a |
| 24         | 239.5 | 8.4   | 511   | 1      | SCRB_ZYMMO | P22632 zymomonas m |
| 25         | 238.5 | 8.4   | 432   | 1      | BRAA_THEMA | Q33833 thermotoga  |
| 26         | 237.5 | 8.3   | 494   | 1      | SCRB_STAXY | Q05936 staphylococ |
| 27         | 232.5 | 8.2   | 484   | 1      | SCRB_VIBAL | P13394 vibrio algi |
| 28         | 222   | 7.8   | 454   | 1      | SCRB_STRMU | P13522 streptococ  |
| 29         | 214   | 7.5   | 466   | 1      | SCRB_SALTY | P37075 salmonella  |
| 30         | 213   | 7.5   | 473   | 1      | SCRB_LACIA | Q04937 lactococcus |
| 31         | 206   | 7.2   | 501   | 1      | SCRB_PEDPE | P43471 pediococcus |
| 32         | 204.5 | 7.2   | 583   | 1      | INV3_DAUCA | Q39693 daucus caro |
| 33         | 197.5 | 6.9   | 480   | 1      | SCRB_BACSU | P07819 bacillus su |

|    |       |     |      |   |            |                    |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 187.5 | 6.6 | 592  | 1 | INV2_DAUCA | Q39692 daucus caro |
| 35 | 125   | 4.4 | 1120 | 1 | STFR_ECOLI | P76072 escherichia |
| 36 | 122   | 4.3 | 3491 | 1 | ERV1_SACER | Q03131 saccharopol |
| 37 | 117   | 4.1 | 451  | 1 | Y483_MVCTU | Q11149 mycobacteri |
| 38 | 116.5 | 4.1 | 984  | 1 | EPA3_RAT   | O08680 rattus norv |
| 39 | 114.5 | 4.0 | 890  | 1 | NTA2_PHAVU | P39866 phaseolus v |
| 40 | 114   | 4.0 | 1132 | 1 | VHSJ_LAMB  | P03749 bacterioph  |
| 41 | 113.5 | 4.0 | 909  | 1 | NIA_PETHY  | P36859 petunia hyb |
| 42 | 113   | 4.0 | 296  | 1 | ALYS_BPDP1 | O03979 bacterioph  |
| 43 | 112   | 3.9 | 916  | 1 | NIA1_ORYSA | P16081 oryza sativ |
| 44 | 111.5 | 3.9 | 752  | 1 | HISA_PROAC | Q59634 propionibac |
| 45 | 110.5 | 3.9 | 983  | 1 | EPA3_HUMAN | P29320 homo sapien |

#### ALIGNMENTS

```
RESULT 1
SACC_BACSU
ID SACC_BACSU STANDARD; PRT; 677 AA.
AC P05656;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan
DE fructanohydrolase).
GN SACC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87286401; PubMed=3112519;
RA Martin I., Debarbouille M., Ferrari E., Klier A., Rapoport G.;
RT "Characterization of the levanase gene of Bacillus subtilis which
RT shows homology to yeast invertase.";
RL Mol. Gen. Genet. 208:177-184(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=88067786; PubMed=3120151;
RA Schoengenderfer K., Schwab H., Lafferty R.M.;
RT "Nucleotide sequence of a cloned 2.5 kb PstI-EcoRI Bacillus subtilis
RT DNA fragment coding for levanase.";
RL Nucleic Acids Res. 15:9606-9606(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=97286548; PubMed=9141695;
RA Parro V., San Roman M., Galindo I., Purnelle B., Bolotin A.,
RA Sorokin A., Mellado R.P.;
RT "A 23911 bp region of the Bacillus subtilis genome comprising genes
RT located upstream and downstream of the lev operon.";
RL Microbiology 143:1321-1326(1997).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=90355183; PubMed=2117666;
RA Martin-Verstraete I., Debarbouille M., Klier A., Rapoport G.;
RT "Levanase operon of Bacillus subtilis includes a fructose-specific
RT phosphotransferase system regulating the expression of the operon.";
J. Mol. Biol. 214:657-671(1990).
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 2,6-beta-D-
CC fructofuranosidic linkages in 2,6-beta-D-fructans (levans)
CC containing more than 3 fructose units.
CC -!- INDUCTION: IN CONTRAST WITH LEVANUCRASE AND SUCRASE, LEVANASE
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE: NO INDUCER IS KNOWN FOR
CC LEVANASE.
CC -!- MISCELLANEOUS: LEVANASE CANNOT BE DETECTED IN THE WILD-TYPE
CC B.SUBTILIS BUT IS MOSTLY SECRETED INTO THE CULTURE MEDIUM BY SACL
CC MUTANTS, ESPECIALLY AT THE END OF THE EXPONENTIAL GROWTH PHASE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05649; CAA29137.1; ALT_INIT.
DR EMBL; Y00485; CAA68542.1; -.
DR EMBL; X92868; CAA63465.1; -.
DR EMBL; X56098; CAA39581.1; -.
DR EMBL; Z99117; CAB4645.1; -.
DR PIR; S06353; S06353.
DR PIR; A27286; A27286.
DR Subtilist; BG10320; sacC.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Signal; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 677 LEVANASE.
FT ACT_SITE 49 49 BY SIMILARITY.
FT CONFLICT 658 658 O -> L (IN REF. 2).
FT SEQUENCE 677 AA; 75951 MW; 80PD6B0A5EE7F525 CRC64;
Query Match 19.5%; Score 556; DB 1; Length 677;
Best Local Similarity 32.2%; Pred. No. 1e-32;
Matches 160; Conservative 74; Mismatches 217; Indels 46; Gaps 18;
QY 40 RAYVHTPPSGWGLCDPQRPVTHGAVQLYLYLHSDONN--CPGGWDHASTTGDGVAFFHGT 97
DB 35 RPOYHFTPEANMNDPNMGVYVYAGEYHLFYQHPYGLQWPMHGHVAKDLTWHEH--L 92
QY 98 VMLRPDP--FPVNSGSAVGTANTAGFGAG---AVVALATQPTDGVKRYQEOVLYKSTDG 152
DB 93 PVALYPDEKTIIFSGSAVDKNTSFGQTEKPLVAIYTDREG---HGVQIAIYNDK 149
QY 153 GFTFTALP-DPVIIVNTDGRAATTPAEIAENAEWFRDPKIHWDTARGEVVCVIGRLVAAFY 211
DB 150 GRWTWKYAGNPVIPN-----PGKKD---FRDPKVFWEKEKKWVMVLAAGDRILLY 197
QY 212 TSNLNRDWTLRNFDYPNHALGI-ECPLDFEITAD--DGRHVLVAASMDAYGIGLPM 268
DB 198 TSNLKNQWTVYASEFGDQSGHGWECPCDLPFLPVCNPNQKKWQVSGVNGAVSGSG 257
QY 269 YAYWTGTWCEQFHADLTPO--WLDGWQWYAAVTPSIDAPETKRLATAWNNNKYAA 326
DB 258 MQYFVGDFDGTHEKNENPPKNKVLWTDYGRDFYAAVSWSDIPSTDRLWLGNWNNWQY-A 316
QY 327 RDVPTDASDYGNGSIVRELRLARPGGWYTLSTPVAAL/TNYVTATTLTLPDRTVDGSA 386
DB 317 NDVPTSP--WRSATSPRELKKAFTG-VRVVQTPVKLEITIRGTSKKWKNLTISPAS 372
QY 387 ---VLPNGRAYEI--ELDTAWDTATNVLISVGRSPDGTHTNIGKYGADLYVDGRPSDL 441
DB 373 HNYLAGSGDAYEINAEFFKVSAAEFKPKVRTGENQFTKVGYDRRNKLFVDRSESGN 432
QY 442 AGYSLA-PYSRAAAPIDPGARSVHLRLVTDQSVQVEFVNAAGH/VLSQQVHFAGDGTGISL 500
DB 433 DTENPAFNTKGTAPLKPNGVKVLRIFVDRSSVEFVNGDQKQVITDIIILPDRSSKGL 492
QY 501 YDGGGAHFTGIVVREI 517
DB 493 YAAANG-----GVRVKSL 504
RESULT 2
FRUA_STRMU
ID FRUA_STRMU STANDARD; PRT; 1423 AA.
AC Q03174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE Fructan beta-fructosidase precursor (EC 3.2.1.80) (Exo-beta-D-
DE fructosidase) (Fructanase).
GN FRUA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=93014178; PubMed=1398976;
RA Burne R.A.; Penders J.E.C.;
RT "Characterization of the Streptococcus mutans GS-5 fruA gene encoding
RT exo-beta-D-fructosidase.";
RL Infect. Immun. 60:4621-4632(1992).
CC -!- FUNCTION: THIS PROTEIN IS A FRUCTANASE ENZYME WHICH DEGRADES
CC LEVANS AND INULINS TO FRUCTOSE AND ALSO CLEAVES SUCROSE INTO
CC GLUCOSE AND FRUCTOSE AND CAN THEREFORE FUNCTION AS AN
CC EXTRACELLULAR INVERTASE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal-non-reducing 2,1- and
CC 2,6-linked beta-D-fructofuranose residues in fructans.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan
CC by an amide bond (potential).
CC -!- INDUCTION: BY SUCROSE, FRUCTAN SUBSTRATES AND FRUCTOSE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U78296; AAA26889.1; -.
DR InterPro; IPR003343; Big_2.
DR InterPro; IPR001362; GH_32.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Pfam; PF02368; Big_2; 1.
DR TIGRFS; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE_NEG.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE_NEG.
KW Hydrolase; Glycosidase; Cell wall; Peptidoglycan-anchor; Signal.
FT SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 1391 FRUCTAN BETA-FRUCTOSIDASE.
FT PROPEP 1392 1423 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 458 458 BY SIMILARITY.
FT DOMAIN 867 871 INVOLVED IN BINDING OF SUGARS WITH
FT BETA-(2,6) LINKAGES OR BINDING OF
FT MOLECULAR WEIGHT FRUCTANS (BY
FT SIMILARITY).
FT SITE 1388 1392 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1391 1391 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT SEQUENCE 1423 AA; 158668 MW; 8E574715F4E72A8A CRC64;
Query Match 13.4%; Score 382.5; DB 1; Length 1423;
Best Local Similarity 28.2%; Pred. No. 8e-20;
Matches 148; Conservative 61; Mismatches 193; Indels 123; Gaps 27;
QY 40 RAYVHTPPSGWGLCDPQRPVTHGAVQLYLYLHSDONNPGGWDHASTTGDGVAFFHGT 98
DB 444 RDQYHYSVKDGMANDPNGLVYNGVHLFHQFYDDTKWGPMMHWAHATSTDLI-----HWKE 499
QY 99 MPLR--PDRP--VWGSVAVGTANTAGF---GAGAVVALATQPTDGVKRYQEOVLYWSTD 151
DB 500 EPIAFYDPSNGYWFSGCVVVDHNSGLFKYAKGLVAITANGNG-----QRMELAYSED 555
QY 152 GGFTFTALPDPIVNTDGRAATTPAEIAENAEW-----FRDPKTI-HWDTARGEVVCV 201
DB 556 EGKTKQKY-DRIV-----ADWSNDPLQNDQFRPKVFHWN---NQFMV 595
QY 202 I--GRLYAAFYTSPNLRDWTLRNFDYPN-HALGGIECPDLFEITADGTRHVLAA 258
```

```

Db 596 LAGGRLP-----YSSNNLKDWKVEST---YDPLHT-----ECPDMIPVANDGVVLKWLSSRG 646
QY 259 DAYGICLPMYAYWTGTWDEGFHADDL---TPQMLDWCWYAAVWT-----PS 305
Db 647 REYKVG---DFKQVDGKW---TFIADDAKDKDQVMFNGKDSYAAWTYYVDFGTETRPT 700
QY 306 IDAPETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELRLARQPCGWTLLSTPVA 365
Db 701 I-----PKTEVNMNTWEDYCNLADVTGQDFNGFNLLNLDGLINE-NGQYILTQTPVK 755
QY 366 ALTNVVTATTL--PDRTVGSAVL--PWNGRAYETEIDIAWDATNNGISVGRSPDGR 421
Db 756 AYDSLRYNTALHFKDVTDVANDANTLLKDFKDSYELVSHFRDEKTT-----802
QY 422 HTNICKYGADLYVDRGSPDSLACYSLAP-----YSRAAAPIDP-----GARS 462
Db 803 -----KVGFNLRVNGQATKVIYDLQTELTSDRSQSGTILSAFAKVNQSHOVTKNADGS 857
QY 463 VHLRLIVDTQSVFVFNAGHTVLSQOVHFAEGDGTISLYTDGGA 507
Db 858 IDLHIYDRASVEVFSKNNTVAGANQIFPNPEAVGASIIIVEGKA 902

RESULT 3
INVL_HANAN
ID INVL_HANAN STANDARD; PRT; 550 AA.
AC P40912;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase).
GN INVL.
OS Hansenula anomala (Yeast) (Candida pelliculosa).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4927;
RN [1]
RP
RQ
RX MEDLINE=96163052; PubMed=8595669;
RZ InterPro: IPR001362; GH_32.
RT Perez J.A., Rodriguez J., Rodriguez L., Ruiz T.;
RT "Cloning and sequence analysis of the invertase gene INV 1 from the
RT yeast Pichia anomala.";
RL Curr. Genet. 29:234-240(1996).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80640; CAA56684.1; -
DR PIR; S48812; S48812.
DR InterPro: IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 550
FT INVERTASE.
FT CARBOHYD 112 112
FT CARBOHYD 113 113
FT CARBOHYD 119 119
FT CARBOHYD 165 165
FT CARBOHYD 211 211
FT CARBOHYD 237 237
FT CARBOHYD 333 333
FT CARBOHYD 364 364

```

```

FT CARBOHYD 398 398
FT CARBOHYD 420 420
SQ SEQUENCE 550 AA; 63188 MW; 1928CD722455E121 CRC64;
Query Match
Best Local Similarity 24.6%; Pred No. 4.5e-20;
Matches 139; Conservative 87; Mismatches 212; Indels 128; Gaps 27;
QY 39 LRAVYHMTPPSGMCLDQRPVTHGAY-----OLYYLHSDNNG-----PGWDHAST 86
Db 27 LRQIHLTPDQGMNDP-----NGMFYDRKDKLWHVYFOHNPDKKSIWATPVTWGHSTS 80
QY 87 TDGVAFTHGTYM-PLRPDPFVWGSASVGTANTAGFGAGA-----VVALATQPTDGYR 139
Db 81 KULLTWDYHGNALPENDEDEGIFSGSVVDNRNNTSGFFNDSTDPEDRIIVATNNA-----136
QY 140 KYQEQLYVSTDGGFTTALPDPIVINTDGRAATTPAEIEENAEWFRDPKIHMDTARGEVW 199
Db 137 QLOTOEIAVSLDKGYSFYIKYDQNPVIN-----NSSQQRDPKVLWHDESQW 184
QY 200 CVIGRLR--YAAFYTPNLRDWTLRNFDPYPNHALGGIECPDLFEITA-----DGTRHW 252
Db 185 MVAKTQEEKVQIYGSPLDKKWLKSNFTSNGYLFQYECPLKLPDENLNDTVTSKW 244
QY 253 VLAASMDAYCIGLPM---TYAYWTGTWDEGFHADDLTPQLWDGWDHWAAYWPSIDAP 309
Db 245 VLLLAIRP---GSLGGISNEYFIDGDTTFHPDDGATRFMDIGKDFAFOSFDNTE-P 300
QY 310 ETKRLAIAWNNWYAAARDVPTDASDGYNGQNSIVRELRLAR---OPGGM-VTLTSTPVA 365
Db 301 EDGALGLANWQY-ANTVPT---ENWRSSMSLVNRYTLKYVDVNPENYGLTLQKPV- 355
QY 366 ALTNVVTATTL-----PDRTVGSAVLPMNGRAYE-----IELDIATWT 405
Db 356 ----YDKETRLNETLKTLETINEYEVNDLKDKSSFATDFNTERNATGVFEFLKF---409
QY 406 ATNVGISVGRSPDGTHTNIGY-----GADLYVDRGP-----438
Db 410 -TOTDLKMGYS---NMTQFGLYIHSQTVKGSOETLQLVFDLTSTTWYIDRTQHSFORN 465
QY 439 SLAGYSLAPYSRAAAPIDPGARSVH-LRLIVDTQSVFVFNAGHTVLSQOVHFAEG--D 495
Db 466 SPVFTERTISTYVEKIDTDDQ--NVYTLGVVDNRNILELYFNDGSIAMNTFFFRGKIP 523
QY 496 TGISLYTDGGAHFT--GIWVREIGQ 519
Db 524 TSFEVVCDSKSFITIDELSVRELAR 549
RESULT 4
INVL_YEAST
ID INVL_YEAST STANDARD; PRT; 532 AA.
AC P10594;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Invertase 1 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 1)
DE (Saccharase).
GN SUC1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP
RQ
RX MEDLINE=88216256; PubMed=2835632;
RT Hohmann S., Gozalbo D.;
RT "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL.";
RL Mol. Gen. Genet. 211:446-454(1988).
RN [2]
RP SEQUENCE FROM N.A.
RQ
RZ Hohmann S.;

```



RL Submitted (NOV-1988) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
CC fructofuranoside residues in beta-D-fructofuranosides.  
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X07570; CAA30457.1; -  
DR PIR: S27372; S27372.  
DR SGD: L0002143; SUC1.  
DR InterPro: IPR001362; GH\_32.  
DR Pfam: PF00251; Glyco\_hydro.32; 1.  
DR PROSITE: PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Hydrolyase; Glycosidase; Glycoprotein; Multigene family; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 532  
FT ACT\_SITE 42 42  
FT CARBOHYD 23 23  
FT CARBOHYD 64 64  
FT CARBOHYD 111 111  
FT CARBOHYD 112 112  
FT CARBOHYD 118 118  
FT CARBOHYD 165 165  
FT CARBOHYD 275 275  
FT CARBOHYD 356 356  
FT CARBOHYD 369 369  
FT CARBOHYD 384 384  
FT CARBOHYD 398 398  
FT CARBOHYD 512 512  
FT CARBOHYD 532 AA; 3280EP260EA01606 CRC64;  
SQ SEQUENCE 532 AA; 60570 MW; 3280EP260EA01606 CRC64;  
  
Query Match 13.0%; Score 369.5; DB 1; Length 532;  
Best Local Similarity 25.1%; Pred. No. 2.1e-19;  
Matches 140; Conservative 84; Mismatches 221; Indels 113; Gaps 25;  
  
QY 19 LALIFGAVPPAASGLRAVYHMTTPSGWLCDPQR--PVTHGAYOLYVHSDQNN 76  
DB 7 LFLAAGFAKISASWTNETSDRLVFTNPKGMNDPNGLWDKAGKWHLFQYNPNDT 66  
  
QY 77 G---PGWDHASTDGCVAATHHG-----TVMPLRPDPVWMSGAVGTANTAGGAGV-- 127  
DB 67 VMGLPLFWGHATSDD---LTHWDEPVAIAPKRKDSGAYSGSMVIDYNNNTSGFFNDITDP 123  
  
QY 128 ---VALATQPTDGVKRYQEOYLWSTDGGFTFTAL-PDPVIVNTDGRAATTPAEIENAE 182  
DB 124 RQRCVAIWYNTP---ESEQYISYSLDGGYTFTEYQKNPVLA-----ANST 167  
  
QY 183 WFRDPKHWDTARGEVWCVIGRLR--YAAFYTSPLNRDWTLRNFDYPNHALGG--IECP 238  
DB 168 QFRDPKVFVYEPKSKWIMTAANKSODYKIEIYSDLLKSKWLESAP--ANEGFLGYQYEC 225  
  
QY 239 DLFEITA--DQCTRHVLWLAASMDAYGIGLPMTYA---YWTGTWGDGQFQHADDLTPOWLDW 293  
DB 226 GLIEVPEQDPKSHWVFMFISINP---GAPAGSGFQYFVSGFNGHFEAFEDNQSRVDF 282  
  
QY 294 GWDYAAVTPSIDAPETRLAIAWNNKYYAARDVPTDASDGYNCQNSIVRELRLAROP 353  
DB 283 GKDYALQTFNTDPTGYGALGIAWNSWEYSA-FYP---SNPWRSSMLSLVPFSLNTE- 337  
  
QY 354 GGNVYLLSTPVAALTNYVATTTLTDPRTVDGSAVLPPN-----GRAYEIEDIAW 403  
DB 338 ---YQANPETELINLKA-----EPILNISSAGPWSRFATNTLTKANSVNVDLSNST 386  
  
QY 404 DT-----ATNVGISVGRS-----PDGTRHTNIGKYADLYVDRGFSDLA 442  
DB 387 GTLEFELVAVVNTQTISKVSFADLSLWPKGLDEPPEYLRMGFEVSASSFSFELDRGNKVK 446

QY 443 GYSLAPY--SRAAAPIDPGARSVHLRI-----LVQTQSVFVFNAGHTVLSQQVFAEGD 495  
DB 447 FVRENPFYFNRMSVNNQPKSENDLSYKYVYGLLDQMLELYENDGDVDVSTNTYFTMTGN 506  
  
QY 496 -----TCIS--LYTD 503  
DB 507 ALGSVNNVTGVDNLFYID 524  
  
RESULT 5  
INVA\_YEAST STANDARD; PRT; 532 AA.  
ID INVA\_YEAST  
AC P10596;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-MAR-1992 (Rel. 21, Last annotation update)  
DE Invertase 4 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 4)  
DE (Saccharase).  
DE SUC4.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE OF 1-73 FROM N.A.  
RX MEDLINE=88216256; PubMed=2835632;  
RA Hohmann S.; Gozalbo D.;  
RT "Structural analysis of the 5' regions of yeast SUC genes revealed  
RT analogous palindromes in SUC, MAL and GAL";  
RL Mol. Gen. Genet. 211:446-454(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hohmann S.;  
RL Submitted (NOV-1988) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
CC fructofuranoside residues in beta-D-fructofuranosides.  
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X07572; CAA30459.1; -  
DR PIR: S27373; S27373.  
DR SGD: L0002146; SUC4.  
DR InterPro: IPR001362; GH\_32.  
DR Pfam: PF00251; Glyco\_hydro.32; 1.  
DR PROSITE: PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Hydrolyase; Glycosidase; Glycoprotein; Multigene family; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 532  
FT ACT\_SITE 42 42  
FT CARBOHYD 23 23  
FT CARBOHYD 64 64  
FT CARBOHYD 97 97  
FT CARBOHYD 111 111  
FT CARBOHYD 112 112  
FT CARBOHYD 118 118  
FT CARBOHYD 165 165  
FT CARBOHYD 256 256  
FT CARBOHYD 275 275  
FT CARBOHYD 356 356  
FT CARBOHYD 369 369  
FT CARBOHYD 384 384  
FT CARBOHYD 398 398  
FT CARBOHYD 512 512  
FT CARBOHYD 532 AA; 60575 MW; A56EA2E3A49125EB CRC64;  
SQ SEQUENCE 532 AA; 60575 MW; A56EA2E3A49125EB CRC64;  
  
Query Match 12.9%; Score 366.5; DB 1; Length 532;







|    |     |  |     |
|----|-----|--|-----|
| Db | 63  | PNDTVGCTPLFWGHAYSDDDLTWNEDQPIAIPARNDSGAFSGSMVVDYNYNTSGFFNDTID  | 122 |
| Qy | 128 | -----VALATQPTDGVKRYQEOYLWSTDGGFTTAL-PPDIVNTDGRAATTAPABEINA     | 181 |
| Db | 123 | PRQRCVAIWTNP---ESEQIYSLDGGYTFTEYQKNPVL-----ANS                 | 166 |
| Qy | 182 | ENFRDPKIHWDTRARGEWCWVIGBLR--YAAFTYSPNLRDWTLRRNDYPNHALGG--IEC   | 237 |
| Db | 167 | TQFRDPKFVFEPSQKWIMTAARKSODYKIELYSDDLKSWKLESAP--ANEGFLGYQYEC    | 224 |
| Qy | 238 | PDLFEITA---DDCTRHWLAAASMDAYIGLPMYIA---YWTGTWDEQFHADDLTPOWL     | 292 |
| Db | 225 | PGLEIVPTEDPSKSYWMEPISINP---GAPAGSGENQYFVGSFNGTHEAFEDNOSRVD     | 281 |
| Qy | 293 | WGWDWTAAVTWPISIDAPETKRLAIAWMNNKYAARDVPTDASDGYNGQNSIVRELRL---   | 349 |
| Db | 282 | FGKDYALQTFEPTDPTYGSALGIAWAKWEYSA-FVPTNP---WRSSMSLVKRTSLNTE     | 337 |
| Qy | 350 | --ARQPGCWYLLSTPVAALTN-----YVATTTLTLDRT--VDGSAVLFWNGRAYEIBL     | 399 |
| Db | 338 | YQANPETELINLKAEPILNISNAGPWSRFATNTTLKANSYNVD-----LSNSTGTLEPEL   | 393 |
| Qy | 400 | DIADWATATNGISVG-----RSPDGRTHNIGKYCADLYVDGRGSDLAGYSLADY         | 449 |
| Db | 394 | YVAVNTQTITSKSVFADLSLWFKGLEDPPEYLRMGFEVSGASSFFLDRGNSKVKYFKVENDY | 453 |
| Qy | 450 | --SRAAAPIDPGARSVHLRT-----LVDYTSQSEVFVNAGHTVLSQQQVHFAEGD---     | 495 |
| Db | 454 | FTNRMSVNNQPFKSENDSLYKKYVGLLDQNLILELYEDGDGVYSTNTYFMTTGNALGSVM   | 513 |
| Qy | 496 | -TGIS--LYTD  | 503 |
| Db | 514 | TTGVDNLFYID  | 524 |

## RESULT. T 8

```

CSCA_ECOLI STANDARD; PRT; 477 AA.
ID CSCA_ECOLI
AC P40714;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sucrose-6-phosphate hydrolase (EC 3.2.1.26) (Sucrase) (Invertase).
DN
GN
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP
RC STRAIN=EC3132;
RC Bockmann J.;
RL Submitted (SEP-1994), to the EMBL/GenBank/DBJ databases.
CC -I FUNCTION: THIS ENZYME ENABLES THE BACTERIA TO METABOLIZE SUCROSE
CC AS SOLE CARBON SOURCE.
CC -I CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -I PATHWAY: Sucrose metabolism.
CC -I SUBCELLULAR LOCATION: Cytoplasmic.
CC -I SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81461; CAA57219.1; -
CC InterPro; IPR001362; GH_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC

```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltyens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Beyer K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880(2002).  
 RN [3]  
 RP SEQUENCE OF 88-581 FROM N.A.  
 RC STRAIN=PR745;  
 RX MEDLINE=98162722; PubMed=950191;  
 RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in Schizosaccharomyces pombe  
 RT cDNAs";  
 RL DNA Res. 4:363-369(1997).  
 RN [4]  
 RP CARBOHYDRATES  
 RX MEDLINE=90253381; PubMed=2187435;  
 RA Moreno S., Sanchez Y., Rodriguez L.;  
 RT "Purification and characterization of the invertase from  
 RT Schizosaccharomyces pombe";  
 RL Biochem. J. 267:697-702(1990).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
 CC fructofuranoside residues in beta-D-fructofuranosides.  
 CC -1- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS 67%. THIS IS  
 CC COMPOSED OF EQUI MOLAR AMOUNTS OF MANNOSE AND GALACTOSE. THERE IS  
 CC ALSO A SMALL AMOUNT OF GLUCOSAMINE PRESENT.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB011433; BAA25684.1; --  
 DR EMBL; AL049644; CAB41057.1; --  
 DR EMBL; D89242; BAA13903.1; --  
 DR InterPro; IPR001362; GH\_32.  
 DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
 DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
 KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 581  
 FT ACT\_SITE 97 97 BY SIMILARITY.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CONFLICT 186 189 MISSING (IN REF. 3).  
 FT CONFLICT 190 190 H -> Y (IN REF. 3).  
 FT CONFLICT 195 195 A -> P (IN REF. 3).  
 FT CONFLICT 198 198 Q -> L (IN REF. 3).  
 FT CONFLICT 205 205 D -> N (IN REF. 3).  
 FT CONFLICT 243 243 I -> M (IN REF. 3).  
 FT CONFLICT 246 246 M -> L (IN REF. 3).  
 FT CONFLICT 256 256 S -> P (IN REF. 3).  
 SQ SEQUENCE 581 AA; 64407 MW; 70206A6CD1F27EC4 CRC64;

Query Match 12.2%; Score 348; DB 1; Length 581;

Best Local Similarity 24.9%; Pred. No. 8.3e-18;

Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;

QY 28 PPAARSAFCSL-----RAYHMTPPSGWLCDDQRPVTHCAQLYY 69

DB 53 PPEVNTTAPNGTCLGNYNEPLSGYINATDRPKIHFTPPSGFMNDPGLYTGCVHMF 112

QY 70 LHSQDQ--NNGPGGWDHASTDTGVAF-----THHGTVMPRLRDPFVWSSGAVVGTGA 117

DB 113 QYSPKTLTAGEVHNGHTVSKDLHWNYPITAIYPDEHENGVLSL-----PFSSAVVDVH 167

QY 118 NTAGFCAG-----AVVALATQPTDGVRYQEQYLXWSTDGGFTTALPDVINYTDGRA 171

DB 168 NSSGLFSNDTIPERIVLIYTDHWTGVA--ERQAIATVTDGGYTFK-----KY 213

QY 172 ATPPAEINAEWFRDPKIHWDTARGWVCVIGRLRY--AAFYISPNLRDWTLRNFDYPN 229

DB 214 SGNPVLIDINSLOFRPKVWDFDANRWMTIVAMSONYAFYSYDLIHWTELSVSTSG 273

QY 230 HALGGTECPDLFEITADDGTRH--WLAASMDAYGIGLPM---TYAYWTGWDGEGQHADD 285

DB 274 YGLQVECEGMARVPVEGDEYKWLFIINP---CAPLGGSVQYFVGDMNGTNFVPDD 330

QY 286 LTPQWLDMGWDVYAAVTPSIDAPETKRLAIAMNNKWAARDVPTDASGYNGQNSIVR 345

DB 331 GQTRFVDLKGDFVASALYHSSA--NADVIGVGWASNNQY--TNOAPTQV---FRSAMTVAR 385

QY 346 ELRLAROPGGWYT---LLSTP--VAALTN--YVATTTLTLDRTVDGSAVLPWNGRAYE- 396

DB 386 KFTLRDVPQNPMTNLTSLQTPLNVLRLRDETFTAPVINSSSSLSGSPITLPSNTAFEF 445

QY 397 -FELDIAMDTATNVGISVGR---SPDGTR---HTNIGKYGADLYVDGRGPSDLGY-SL 446

DB 446 NVTLISNYEGTGTGCLGRIIISDDPYRLQSLISVDVDFAASTLVINRAKAKMGWENSL 505

QY 447 APYSRAAAPIDPGARSVHLRLVDTSQSVFVFNAGHTVLSQQVHFAEGDT 496

DB 506 FTSPFANDIYIYG--NVTLYGIVDGLLELYVNGEKTNTDFFFLQGT 553

RESULT 10

RAFT\_ECOLI

ID RAFT\_ECOLI

AC P16553;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-MAR-1992 (Rel. 21, Last annotation update)

DE Raffinose invertase (EC 3.2.1.26) (Invertase).

GN RAFT.

OS Escherichia coli.

OG Plasmid pRSD2.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.



```
Db 173 -SSNQPRDPKVFHWRPKSMHGCSE-----IARVKIQIF-GSANLKNWVLSNFSFGYYG 226
QY 230 HALGGIECPDLFEITADGTR-HWVLAASMDAYGICGLPM---TYAYWTGTWDEQFHADD 285
Db 227 NOYG---MSRLIEVPIENSCKWVFLAIPN---GSPLGSSINQYFVGDFDGFQFVPDD 280
QY 286 LTPQWLDWGDWTAANTWPSIDAPETKRLAIAWMNNKYYAARDVPTDASGYNGQNSIVR 345
Db 281 SOTRFVDIGDKDFAFQTFSEV---EHGVLGLAWASWQYADQ-VPTNP---WRSSTSLAR 333
QY 346 ELRLARQPGGWTLLSTPVA---LTPNYTATITLDPRTDGGSAVLP-----WNGRAY 395
Db 334 NYTLR-----YVIQMLKUTANIDKSVLPDSINVVDKLLKNKVLTKNPKFNGSGTG 387
QY 396 EIELDIAWDATATNVGISVGR-----SPDGRTRHNIQIGYAD-----LYVDRG----- 437
Db 388 LFDENITF-KVLNLNVSPCKTHFDILINSQELNSSVDSIKIGFSSQSLSFYIDRHNPVE 446
QY 438 -----PSDLACYSLAPYSRAAAPDPGARSVHLRLVDVTSQVEFVNAGHTVLSQQVH 490
Db 447 FPRKQFFTDKLAAY-LEPLD-----YDQDLRVESLYGIVDKNIIELFYNDGTVMANTTF 500
QY 491 FAEG-----DTGISLYTDGGAHFTGTIVWREIGQ 519
Db 501 MGEKYPHDIQIVTDTEEPLEFLESVIIRLKN 533

RESULT 12
INVA_MAIZE
ID INVA_MAIZE STANDARD; PRT; 590 AA.
AC P49174;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Beta-fructofuranosidase, cell wall isozyme precursor (EC 3.2.1.26)
DE (Sucrose-6-phosphate hydrolase) (Invertase).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Black Mexican Sweet;
RA Shanker S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -! SUBCELLULAR LOCATION: Cell wall.
CC -! INDUCTION: BY wounding and bacterial infection.
CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U17695; AA064487.1; -
CC MaltedB; I13032; -
CC InterPro; IPR001362; GH_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE NEG.
KW Hydrolase; Glycosidase; Glycoprotein; Cell wall; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 590 BETA-FRUCTOFURANOSIDASE, CELL WALL
FT ACT_SITE 68 68 ISOZYME.
FT CARBOHYD 190 190 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 590 AA; 6519 MW; 296455E61E9B892 CRC64;
```

```
Query Match 11.0%; Score 312; DB 1; Length 590;
Best Local Similarity 25.3%; Pred. No. 3.3e-15;
Matches 150; Conservative 81; Mismatches 211; Indels 150; Gaps 34;

QY 19 LALIFGA-----VPPAARASAPGS-----LRVYHMTTPSGWLCDDPQRPVTHCAYQ 66
Db 21 LALLRAGASHVHRSLEAEAPSPASIVSPLRLTGYHFQPPMNNWINDPNAPLYYKGYH 80
QY 67 LYVLHSDONNGPG--WDHASTTGDVAFTH-HGTVMPLRPD--FPVWMSGSAVGTANTAG 121
Db 81 LFTQYKPGKAVGNIVVAHSVSRDLINWVALEPAIYPSIPSDKYGCWSGATI-----LE 135
QY 122 FGAGAVVALATQDTPGVRYQEQLYXWSTDGGFTTALP-----DPVIVNTD-----GRAA 172
Db 136 DGTPAIYLTGIDRAD--INVQVQL-----ALPKDASDPLLRWEKPEEYNPVA 182
QY 173 TTPAEIENAEWFRDPKIHWDTARGEVVCVIGRLR-----YAAFTYSPMLRDMWTLRRNEDYP 228
Db 183 TPAAGGINATQFRDPTTAMRHA-GHWRMLVSGVRGARGMALVYRSRDRFRKWKAK---HP 238
QY 229 NH--ALGGI-ECPLFEI-----TADGCTRHVWLAASMDAYGIGLPMYVAYWT-G 274
Db 239 LHSAAITGMWECDFPVPVSGPGLQAGLDTSAPGK-YVLKSSLDL-----TRYDYITIG 291
QY 275 TWDG--EQFHADDLTPQW-----LDMGWDAAYAAVTWPSIDAPETKRLAIAWMNNKYYAAR 327
Db 292 SYDGGKDRYYPDPAGDYHRRRYDYG-NYYASKTF--YDPVERRRVLLGWANE---SD 344
QY 328 DVPTDASDNGYNGNSIVRELRLARQGGWYTLSTPVAALTNVYVATTTLPDRTVDGSAV 387
Db 345 SVPDCKAKAGWAGIHAIPRKIWL--DPTG-KQLLOWPIHEVEKLRGKAV-----SVDKLV 396
QY 388 LPWNG-----RAYETLDTAWDTATNVGISV-----GRSPDGTHTNIGYKADLYV 434
Db 397 KPGDHFVETGIATYQADVEVSFELEAGTSLLEKAEAFDPAYDDDAQKICGVKGADARG 456
QY 435 DRGP-----SDLAGYSLAP---YSRAAA---P 455
Db 457 GVPFGGLWLASADLOERTAVFVRFDGHHGKPKVLMCTDPTKSSLSPLDYKPTFAFVD 516
QY 456 IDPGARSVHLRLVDTQSVSEFVNAGHTVLSQQVH--FAEG-DTGISLYTDG 504
Db 517 ADISSGKITLRSIDRSVSVESFGAGGKTCILSRVYPSIAVGKDAHLVYFVNG 568
```

```
RESULT 13
INVA_ZYMO
ID INVA_ZYMO STANDARD; PRT; 512 AA.
AC P35636;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Sucrose-6-phosphate hydrolase E1 (EC 3.2.1.26) (Sucrase E1)
DE (Invertase E1).
DE INVA.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Z6C;
RA MEDLINE=91291345; PubMed=1368686;
RA Yanase H., Fukushi H., Ueda N., Maeda Y., Toyoda A., Tomomura K.;
RT "Cloning, sequencing, and characterization of the intracellular
RT invertase gene from Zymomonas mobilis.";
RL Agric. Biol. Chem. 55:1383-1390(1991).
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
```





```
Db 65 SPAPPSRGVSGVSDKTRFDVAGASHVSYAWSNMLSWQRTAYHFQPKQNMNDPGLY 124
QY 61 THGAYLYLHSDQNN--GPGGWDHASTTGDGVAFTHHGTVM--LRPD-----FPVWSGS 111
Db 125 HKGWYHLFYQNPDSAIWNTGHAHSVLDLI-----HWLYLPFAMVPDQWYDINGVWTGS 180
QY 112 AVVGTANTAGFAGAGAVVALATOPTDGVRYQEQYLYXWSTDGFTTALPDPIV---VNTD 168
Db 181 ATT-----LPDGLIMLYTGTDDY--VQVONLAYPAN-----LSDPLLLDWVKF 224
QY 169 GRAATTPAEIAENAEWRDPKIH--DTARGEWCV-----IGRLRYAAFYTSNLRDWTLLR 223
Db 225 GNPVLPPPGIGVKDRDPTTAWTGPQNGQWLLTIGSKICKTGVALVYETSFTSKL-- 282
QY 224 NFDPYPHALGGI---ECPDLFEITA-----DDGTRHWVLAASM-----DAYGTG 264
Db 283 -LDGVLHVPVPGTGWECVDFYVSTKTKTNGLDTSYNGPGVKH--VLKASLDDNNKQDHYAIG 340
QY 265 LPMTYAYWTGTWDEQFHADDLTPQWLDGWDWYAAVTPSIDAPETKRLALAMNMWKY 324
Db 341 ---TYDLGKNKTPDNPDLDCGIGLRLDYG--KYASKTF--YDPKKERRVLGWIGETDS 394
QY 325 AARDVPTDASDYGNGNSIVRELRLAROPCGWYTLTSTPVAALTNVVTATTLTPDRTVD- 383
Db 395 ESADL-----OKGWSVQSIPTVLYDKKTG--THLLQWVPEEIESLURVGDPVT--KQVDL 446
QY 384 -GSAPLPWNGRAYEIEIDAMDT-----ATNUGISVGRSPDGTGTRHNTIGKYG-- 429
Db 447 OPGSIELLRVDSAAELDIASFEVDKVALOGIIEADHVGFCSTSGAASRGILGPGVI 506
QY 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAIPDPGARS 462
Db 507 VIADQTLSELTPVYFISKADGAEHFCADQTRSEAPGVGKQVYGSVPVLDCEKHS 566
QY 463 VHLRLVDTQSVFVFNAGHTVLSQOVH 490
Db 567 --MRLLVDSHIVESFAOGGRTVITSRIY 592
```

## RESULT 15

```
INVL_CAPAN STANDARD; PRT; 640 AA.
AC P93761.1
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Acid beta-fructofuranosidase AIV-18 (EC 3.2.1.26) (Acid sucrose-6-
DE phosphate hydrolase) (Acid invertase).
OS Capsicum annuum (bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fruit;
RA Choi D., Lee K.-W., Kim S.;
RT "Isolation and characterization of acid invertase cDNA clone in Hot
RT pepper (Capsicum annuum L.) fruits.";
RL J. Plant Biol. 40:298-303(1997).
CC -/- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -/- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL: U87849; AAB48484.1; -.
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
DR PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Transmembrane; Glycoprotein.
FT TRANSMEM 34 54 POTENTIAL.
FT ACT_SITE 126 126 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 612 612 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 640 AA; 70620 MW; D3C628B7A7E6870B CRC64;

Query Match 9.2%; Score 260.5; DB 1; Length 640;
Best Local Similarity 23.8%; Pred. No. 1.9e-11;
Matches 138; Conservative 77; Mismatches 221; Indels 143; Gaps 32;

QY 1 MTPAISRRVLAQAGAGALALIFGGAVPPAARASAP--GSLRAYVHMTPPSGWCLDQRPV 59
Db 72 LTPATPSRGVSQGVSEKTFKDSVTSQSVTSNAMLNMQRTAYHFQPKQNMNDPGL 131
QY 60 LTHGAYOLYLYLHSDQNN--GPGGWDHASTTGDGVAFTHHGTVM--LRPD-----FPVWSG 110
Db 132 YHKGWTHLYFYQNPDSAIWNTGHAHVSTDLI-----HWLYLPFAMVPDQWYDINGVWTG 187
QY 111 SAVVGTANTAGFAGAGAVVALATOPTDGVRYQEQYLYXWSTDGFTTALPDPIV---VNT 167
Db 188 SATI-----LPDGLIMLYTGTDDY--VQVONLAYPAN-----LSDPLLLDWVKY 231
QY 168 DGRAATTPEIAENAEWRDPKIH--DTARGEWCV-----IGRLRYAAFYTSN--LRDWT 220
Db 232 QGNPVLPPPGIGVKDRDPTTAWTGPQNGQWLLTIGSKVGTGIALVYETSFKLLDGV 291
QY 221 LRRNFDPYPHALGGI---ECPDLFEITAADD-----GTRHWVLAASMDAYGIGLP 266
Db 292 L-----HAVPGTGWECVDFYVSTLDANGLDTSYNGPGIKH--VLKASLDD-----N 337
QY 267 MTYAYWTGTWD--GEOFHADDLTPQW-----LDKGWDWYAAVTPSIDAPETKRLAIA 317
Db 338 KQDHYVIGTVDPVKNKFSQDN--PDLDGCGIGLRLDYG--RYASKTF--YDPKKQRRVLWG 392
QY 318 WNNWKYAADVPDTPDASDYGNGNSIVRELRLAROPCGWYTLTSTPVAALTNVVTATTL 377
Db 393 WIGETDSESADL-----OKGWSVQSIPTVLYFDDKTG--THLLQWVPEEIESLURSGDPK 446
QY 378 PDRT-----VDGSAVLPWNGRAYEIEIDAMDTAT-----NVGISVGRSPDGT 420
Db 447 KEVNLOPGSIELLHVDSSA-----QFDEASFEDRVTLLEGIIADYGYNCSTSGGAA 499
QY 421 RHTNIGKYG-----AD-----LYVDRG-----PSDLAGYSLAPYSR 451
Db 500 SRGILGPGFVGVVIAQDTLSELTPVYFISRGADGRAEAFHFCADQTRSEAPGVAKQVYG- 558
QY 452 AAAPIDPGARSVHLRLVDTQSVFVFNAGHTVLSQOVH 490
Db 559 SSVFVLDEGKH--RMRLLVDSHIVESFAOGGRTVITSRIY 596
```

Search completed: May 6, 2003, 12:32:39

Job time : 14.5202 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:31:02 ; Search time 19.5375 seconds  
(without alignments)  
784.611 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845

Sequence: 1 MTPAIRRAVLQAGAGALA.....TDGGPAHFTGIVVREIQGAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 499.5 | 17.6        | 943    | 4  | US-09-397-885-5     |
| 2          | 488.5 | 16.5        | 1277   | 4  | US-09-397-885-3     |
| 3          | 461   | 16.2        | 923    | 4  | US-09-397-885-1     |
| 4          | 348   | 12.2        | 581    | 4  | US-09-331-581-2     |
| 5          | 333.5 | 11.7        | 337    | 4  | US-09-331-581-24    |
| 6          | 291   | 10.2        | 332    | 4  | US-09-331-581-23    |
| 7          | 275   | 9.7         | 635    | 1  | US-08-245-809-1     |
| 8          | 274.5 | 9.6         | 636    | 1  | US-08-296-624-2     |
| 9          | 235.5 | 8.3         | 492    | 4  | US-09-134-001C-3895 |
| 10         | 205.5 | 7.2         | 630    | 3  | US-08-860-091A-2    |
| 11         | 205.5 | 7.2         | 630    | 4  | US-09-245-323A-6    |
| 12         | 189   | 6.6         | 615    | 3  | US-08-860-091A-4    |
| 13         | 189   | 6.6         | 615    | 4  | US-09-245-323A-8    |
| 14         | 169.5 | 6.0         | 626    | 4  | US-09-019-385-2     |
| 15         | 147.5 | 5.2         | 565    | 4  | US-09-142-623-11    |
| 16         | 141.5 | 5.0         | 635    | 4  | US-09-142-623-1     |
| 17         | 140.5 | 4.9         | 574    | 4  | US-09-142-623-13    |
| 18         | 122   | 4.3         | 3491   | 2  | US-07-642-734C-2    |
| 19         | 122   | 4.3         | 3491   | 3  | US-08-439-009A-2    |
| 20         | 119.5 | 4.2         | 3739   | 3  | US-09-320-878-2     |
| 21         | 115   | 4.0         | 3739   | 4  | US-09-105-537-33    |
| 22         | 115   | 4.0         | 11877  | 4  | US-09-105-537-6     |
| 23         | 110.5 | 3.9         | 983    | 1  | US-08-167-919A-10   |
| 24         | 110.5 | 3.9         | 983    | 2  | US-08-449-645A-21   |
| 25         | 110.5 | 3.9         | 983    | 2  | US-08-702-367A-21   |
| 26         | 110.5 | 3.9         | 983    | 3  | US-08-715-106-10    |
| 27         | 110.5 | 3.9         | 983    | 5  | PCT-US95-04681-21   |

|    |       |     |      |   |                   |                   |
|----|-------|-----|------|---|-------------------|-------------------|
| 28 | 106   | 3.7 | 3519 | 4 | US-09-428-517-4   | Sequence 4, Appli |
| 29 | 106   | 3.7 | 4551 | 3 | US-09-320-878-1   | Sequence 1, Appli |
| 30 | 106   | 3.7 | 4613 | 4 | US-09-105-537-31  | Sequence 31, Appl |
| 31 | 105.5 | 3.7 | 811  | 1 | US-08-480-604A-7  | Sequence 7, Appli |
| 32 | 105.5 | 3.7 | 811  | 2 | US-08-405-496A-7  | Sequence 7, Appli |
| 33 | 105.5 | 3.7 | 811  | 4 | US-08-915-136-7   | Sequence 7, Appli |
| 34 | 105.5 | 3.7 | 811  | 4 | US-08-957-310-7   | Sequence 7, Appli |
| 35 | 105.5 | 3.7 | 812  | 1 | US-08-480-604A-29 | Sequence 29, Appl |
| 36 | 105.5 | 3.7 | 812  | 4 | US-08-915-136-29  | Sequence 29, Appl |
| 37 | 105.5 | 3.7 | 2710 | 1 | US-08-480-604A-6  | Sequence 6, Appli |
| 38 | 105.5 | 3.7 | 2710 | 2 | US-08-405-496A-6  | Sequence 6, Appli |
| 39 | 105.5 | 3.7 | 2710 | 4 | US-08-915-136-6   | Sequence 6, Appli |
| 40 | 105.5 | 3.7 | 2710 | 4 | US-08-957-310-6   | Sequence 6, Appli |
| 41 | 104.5 | 3.7 | 964  | 4 | US-08-484-791-2   | Sequence 2, Appli |
| 42 | 104   | 3.7 | 655  | 4 | US-09-311-626B-14 | Sequence 14, Appl |
| 43 | 104   | 3.7 | 3033 | 1 | US-07-925-695-5   | Sequence 5, Appl  |
| 44 | 103   | 3.6 | 936  | 5 | PCT-US94-05905-22 | Sequence 22, Appl |
| 45 | 102.5 | 3.6 | 614  | 1 | US-08-262-338A-4  | Sequence 4, Appli |

ALIGNMENTS

RESULT 1  
US-09-397-885-5  
; Sequence 5, Application US/09397885  
; Patent No. 6323007  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rahbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6-b-D-Fructan Hydrolase Enzyme And  
; TITLE OF INVENTION: Process For Using The Enzyme  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/397,885  
; CURRENT FILING DATE: 1999-09-17  
; EARLIER APPLICATION NUMBER: PA 1998 01173  
; EARLIER FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: PA 1998 01623  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 60/101,615  
; EARLIER FILING DATE: 1998-09-24  
; EARLIER APPLICATION NUMBER: 60/111,675  
; EARLIER FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-09-397-885-5

Query Match 17.6%; Score 499.5; DB 4; Length 943;  
Best Local Similarity 30.8%; Pred. No. 1e-36;  
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;  
QY 40 RAVYHMTSPGMLCDPQRPVTHGAYQLYLLHSDNNGPGWDHASTDDGVAFTHGVTM 99  
DB 383 RPDYHSPARGSADPNGLVYGEYHLF--HQD---GGTWAHAVSTDLVHWKRLPIAL 436  
QY 100 PLRPFDPVWSSGAVGTANTAGF-----CAGAVVALATOPTDGVKRYQOYLWSTDDGF 154  
DB 437 PMNDLGHVWSSGAVADLNHNASGLFADSGKGLIAYTTSYNPDRPNCNQRIGLAYSKDRGR 496  
QY 155 TPT-ALPDPIVINTDGRAATTPAEIENAEW-FRDPCKIHMDTARGEWVCVIGRLRYAAFT 212  
DB 497 TWYEAERPIVIENFGKQDDP-----GGWDFRDKPVVDEEHNKRWVWVSGDHIRFT 551  
QY 213 SPNLRDWTLRNFDYFNHALGGI-ECPLDFEITADD-GTRHWVLAASMDAYGIGLPMTYA 270  
DB 552 STNLDWTLTDSFGYGAVYRGVWECPLDLQLAVDDTGEKWLMTSGAN-----PNTQG 607

QY 271 YMTGTWDEQGFHADDLTPQW-----LDMGWDWYAAVWTWPSIDAPETKRLAIA 317  
Db 608 -----SAAEYFEGELTPEGKFVNDNPAGKVLATDYGKEYASMSFAGM--PDGRRVMLA 659  
QY 318 WNNWYKAAARDVPTDASDYGNGNSIVRELRLARQPGGWYTLSTPVAALTN-----YVTA 373  
Db 660 WMTNNDYPFA-FPT-----EGWKVLSPRELTLOKTDKG-IRLAQTPIRELESIRGOLLEA 714  
QY 374 TTTLPORVDG-----SAVLPWNGRAYEIELDIAWDTATNNGISVGRSPDGTFRHNTICKYGA 430  
Db 715 AS-----DRRVQADRENLLKGVSSGVYIEAEIEIPQASNVSEFGFRLREGAKRTVVGYKT 771  
QY 431 ----DLVDRGSPDLACYSLAPYSRAAAPIDPGARSVHLRLVDTQSVFVNAGHTVLSQ 487  
Db 772 KENEIVDRSLSGDTCFSEFRTTLHQAPLQPNRRVKRLRIFVDDSSLEVFEGDGRVVFSE 831  
QY 488 OVHFAEGDGTISLYTDGGPAHTGIIVREI 517  
Db 832 VIFPDPAHREMSLFTVGGEVNVVSLKVHAL 861

## RESULT 2

US-09-397-885-3  
; Sequence 3, Application US/09397885  
; Patent No. 6323007  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rahbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/397,885  
; CURRENT FILING DATE: 1999-09-17  
; EARLIER APPLICATION NUMBER: PA 1998 01173  
; EARLIER FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: PA 1998 01623  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 60/101,615  
; EARLIER FILING DATE: 1998-09-24  
; EARLIER APPLICATION NUMBER: 60/111,675  
; EARLIER FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1277  
; TYPE: PRT  
; ORGANISM: Paenibacillus pabuli  
US-09-397-885-3

Query Match 16.5%; Score 468.5; DB 4; Length 1277;  
Best Local Similarity 29.1%; Pred. No. 1.1e-33;  
Matches 168; Conservative 74; Mismatches 243; Indels 93; Gaps 25;

QY 4 AISRRAVLQAGA-----GALALIFGGAVPAA-----RASAPG----- 37  
Db 320 ASSNRKVPQPCARHIEVIASSLIQVVDGYTPAAVEVTDKSYAKGNAGLVVQGMAY 379  
QY 38 -----SLRAYHMTPPSGMLCDPQRPVTHGAYQLYYLLHSDQNGGWD 82  
Db 380 FODIYTEESMYKENVYRQYHSPYLRGSASDPNGLVYVEGEYHLF--HOD-----GGTWA 433  
QY 83 HASTDGVAFTHGVTMLRPFVWGSAAVGTANTAGF-----CAGAVVALATOPTDG 137  
Db 434 HAVSSDLINWKLRLPALPNDGCHVWGSAGIADLNNSGLFTSDGKGLIAYTSTHPDK 493  
QY 138 VRKYEQVLYWSTDGGFTFT-ALPDPIVYNTDGRAATTPAEIAENAEW-FRDPKIHWDTAR 195  
Db 494 PGGNQRIGLAYSTDGGRNWOYAKERPIVIDNPGKNGDDP-----GSWDFRDPKVVREDH 548

QY 196 GEMVVCVIGRLRYAAAYTSPNLRDWTLLRRNFDYPNHALGGI-ECPDLEFETAD-DGTRHWV 253  
Db 549 NRWVMVSGDHLRFFTTSTNLLDNTNFYGDYVRGGWVECPDLIQLPVOCTGORKWV 608  
QY 254 LAASMDAYGIGLPMYFA-----YWTG--TWDEGEOFHADDLTPQWL--DWGHDWYAAWTWPS 305  
Db 609 LIISTGAN-----PKTGSDAEYFVGQLTADG-KFLNDHPAGQVLRDIDYKKEFYASMSFAN 663  
QY 306 IDAPETKRLAIAWNNWYKAAARDVPTDASDYGNGNSIVRELRLARQPGGWYTLSTPVA 365  
Db 664 M--PNQRKVLAWMTNNDYFF-EFTSS---WKQGLTIPREVSRLTDEG-VRLVQTPIT 716  
QY 366 ALTNVYVATTLPDRVDSGSAVLPMWG---RAYEIELDI---AWDTATNNGISVGRSPDG 419  
Db 717 ELQKLRHNLVSAQMTVGPKNPLEGLTAGAYEIEAEVEIPANSSTVEFGFOL-RQREG 775  
QY 420 TRHTNIGKYGAD--LYVDRGSPDLACYSLAPYSRAAAPIDPGARSVHLRLVDTQSVFV 476  
Db 776 QKTT--VAYRVDTONMFVDRTTSGDVSFSLFTTKVHEASLKPENQKVLRIFFVDESSEV 833  
QY 477 FVNAGHTVLSOOVHFAEGDGTISLYTDGGPAHTGIW 514  
Db 834 FGNDGKVFESDVIFFDPPAGRAMAFYSLGGEVKSVMKV 871

## RESULT 3

US-09-397-885-1  
; Sequence 1, Application US/09397885  
; Patent No. 6323007  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rahbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/397,885  
; CURRENT FILING DATE: 1999-09-17  
; EARLIER APPLICATION NUMBER: PA 1998 01173  
; EARLIER FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: PA 1998 01623  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 60/101,615  
; EARLIER FILING DATE: 1998-09-24  
; EARLIER APPLICATION NUMBER: 60/111,675  
; EARLIER FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 923  
; TYPE: PRT  
; ORGANISM: Paenibacillus amylolyticus  
US-09-397-885-1

Query Match 16.2%; Score 461; DB 4; Length 923;  
Best Local Similarity 29.5%; Pred. No. 3.1e-33;  
Matches 150; Conservative 74; Mismatches 223; Indels 62; Gaps 23;

QY 40 RAVYHMTPPSGMLCDPQRPVTHGAYQLYYLLHSDQNGGWDHASTTQGVAFTHGVTW 99  
Db 392 RPOYHTPIRGASDPNGLVYVEGEYHLF--HOD-----GGTWAHAYSKDMLNKRLPAL 445  
QY 100 PLRDPFFVWGSAAVGTANTAGF-----CAGAVVALATOPTDGVKRYQBYLYWSTDGGF 154  
Db 446 PWNDCGHVWGSAAVADMTNASGLFGDSGKGLIAYTTFENPSPNCRIGLAYSKDOGR 505  
QY 155 TFTAALPD-PVIVNTDGRAATTPAEIAENAEW-FRDPKIHWDTARGEVVCVIGRLRYAAFT 212  
Db 506 TWEYSKERPIVNIENPKSGN-----EAGNDFRDPKVIIRDENNRRWVMVSGDHLRFFVT 560  
QY 213 SPNLRDWTLLRRNFDYPNHALGGI-ECPDLEFETADGTT--RHWVLAASMDAYGIGLPMY 268

Db 561 STNLLDWTLDNKGVDYVRGWECPDLFQLPV-DGTSQKKVWMISTGAN-----PKTG 615  
Qy 269 ---YAWTQ--TWGDEQFADDLTPOWL--DWGWDYAAVTPSIDAPETKRLAIAMNNK 321  
Db 616 GSDAEYFICHLTADG-KFVNDNPCKVLETDGKKEFYASMSFANM--PDHRTVMMAWMTN 672  
Qy 322 WKYAADVPTDASDGYNGQNSIVRELRLARQPGGWTLLSTPVAALTNVVTATTLPDRT 381  
Db 673 WDYPEA-FPT---SNWKGELTIPREVSLVTTEDG-IRMVQSPKEL-----ESLRKPLYS 722  
Qy 382 VDSAVLPWNGR-----AYEIELDI---AWDTATNVGISVGRSPDGTTRHNTICKYGA 430  
Db 723 ASKSVSPSSGNLLKGIISGAYEIEAEIEIPETSVTEFGFIREG--ANOKTVVGYKAS 780  
Qy 431 D--LVVDRGPSLACYSIAPYRAAAPIDPGARSHLRILVDTQSVFVFNAGHTVLSQO 488  
Db 781 DSRMEVDRTASGETDFSNLFSKKHEAQTOMENNRKMRILVDSESSVEAFNGDKGVVFS 840  
Qy 489 VHAEGDTGISLYTDGGAHFTGIVVREI 517  
Db 841 IPPDPASRAMSFYKGGNVVYSLKVHQL 869

## RESULT 4

US-09-331-581-2

; Sequence 2, Application US/09331581

; Patent No. 6130070

; GENERAL INFORMATION:

; APPLICANT: TOHDA, Hideki

; APPLICANT: KUMAGAI, Hiromichi

; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE

; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS

; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF

; FILE REFERENCE: 0059-1142-0PCT

; CURRENT APPLICATION NUMBER: US/09/331.581

; EARLIER FILING DATE: 1999-06-30

; EARLIER APPLICATION NUMBER: PCT/JP98/04929

; EARLIER FILING DATE: 1998-10-30

; EARLIER APPLICATION NUMBER: JP 9-314608

; EARLIER FILING DATE: 1997-10-31

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 2

; LENGTH: 581

; TYPE: PRT

; ORGANISM: Schizosaccharomyces pombe

US-09-331-581-2

Query Match 12.2%; Score 348; DB 4; Length 581;  
Best Local Similarity 24.9%; Pred. No. 3e-23;  
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;

Qy 28 PPAARASAPGSL-----RAVYMTTPSGWLCDDPQRPVTHGAYQIYV 69  
Db 53 PPVNTTAPNCTCLGNRYNEIPLSGYINATDRPKIHFTSPSGEMDNPGLVVTGGYHMF 112  
Qy 70 LHSDDQ--NNGPGGWDAHSTDGVAF-----THGTVMLRDPDPVNSGSAVVGTA 117  
Db 113 QYSPKTLTAGEVHGHGTVSKDLIHENYPIAIPDEHENGVLUSL-----PFGSAVDVH 167  
Qy 118 NTAGFAG-----AVVALQTPTDGRVKYQBYLYWSTDGGFTFTALPDPIVNTDGRA 171  
Db 168 NSSGLFSNDTIPERIVLIYTDHWTGVA--ERQATAYTDDGGYTPK-----KY 213  
Qy 172 ATTPAEINAEWFRDPKTHWDTARGEWCVCIGRLY--AAFYTSNLRDWTLRNFDYPN 229  
Db 214 SGNPVLIDNSLQRPKPVINDFANDRWVIMVAMSONYGIATFVSSYDLIHWTELSVFSTG 273  
Qy 230 HALGGIECPDLFEITADGTRH-WVLAASMDAYGIGLPM---TYAYWTGTWDGQFHAD 285  
Db 274 YLGQYECPCGMARVPEGTDEYKWLFLISINP---GAPLGSSVVOQYFVGDWNGTNFV 330

Qy 286 LTPQWLMDGWDWYAAVTPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVR 345  
Db 331 GQTRFVLDGKDFYASALYHSSSA-NADVIGVGNASWQY-TNQAPTQV---FRSAMTVAR 385  
Qy 346 EURLARQPGGWT-----LLSTP--VAALTN--YVTTATTLPDRTVDGSAVLDPWNGRAYE- 396  
Db 386 KFTLRDVPQNPMTNLTSLIQTPLNVSLLRDETLFTAPVINSSSSLSGSPITLPSNTAFEF 445  
Qy 397 -TELDIANDTATNVGISVGR---SPDGR---HTNIGKYGADLYVDRGPSDLAGY-SL 446  
Db 446 NVTLSINTEGCTTGYCLGRIIISDDPYRLOSISVDVDFAASTLVINRAKAGMWFNSL 505  
Qy 447 APYRAAAPIDPGARSHLRILVDTQSVFVFNAGHTVLSQOQVHFAEGDT 496  
Db 506 FTFPSANDIYIG--NVTLYGIVDNGLLELYVNGEKTVTYNDFFFLGGAT 553

## RESULT 5

US-09-331-581-24

; Sequence 24, Application US/09331581

; Patent No. 6130070

; GENERAL INFORMATION:

; APPLICANT: TOHDA, Hideki

; APPLICANT: HAMA, Yuko

; APPLICANT: KUMAGAI, Hiromichi

; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE

; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS

; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF

; FILE REFERENCE: 0059-1142-0PCT

; CURRENT APPLICATION NUMBER: US/09/331.581

; EARLIER FILING DATE: 1999-06-30

; EARLIER APPLICATION NUMBER: PCT/JP98/04929

; EARLIER FILING DATE: 1998-10-30

; EARLIER APPLICATION NUMBER: JP 9-314608

; EARLIER FILING DATE: 1997-10-31

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 24

; LENGTH: 337

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-331-581-24

Query Match 11.7%; Score 333.5; DB 4; Length 337;  
Best Local Similarity 30.0%; Pred. No. 2.7e-22;  
Matches 105; Conservative 55; Mismatches 139; Indels 51; Gaps 16;

Qy 24 GGAVPPAARASAPGSLRAVYHMTTPSGWLCDDP-----QRPVTHGAYQIYLYLHSDQNG 77  
Db 10 GPAAKKSAMTNETSDRPLVHFTPNKGMNDPNGLWYDEKDAKWHYFQ--YNPNDTVMG 67  
Qy 78 -PGGDHASTTDGVAFTHHG-TVMPLRDPFPVWGSAAVGTANTAGFCGAGV-----VA 129  
Db 68 TPLFWGHATSDOLTWNEDQPIAIPKRNDSGAFSGMSWVDYNNYNTSGFFNDTIDPQRCVA 127  
Qy 130 LATQPTDGRVKYQBYLYWSTDGGFTFTAL-PDPVIVNTDGRAATTPAEINAEWFRDPK 188  
Db 128 IWTYNTP---ESEQYISYSTDGGYTFTEYQKNPVLA-----ANSTQFROPK 171  
Qy 189 IHWDTARGEWCVCIGRLR--YAAFYTSNLRDWTLRNFDYPNHALGG--IECPDLFEIT 244  
Db 172 VEWYEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKTESAF--ANEGLFGYQYECPCGLIEVP 229  
Qy 245 A--DDGTRHVVLAASMDAYGIGLPMTYA---YWTGTWDGQFHADDLTPQWLMDGWDWYA 299  
Db 230 TEQDPSKSYWWMFISINP---GAPAGGSFNQYFVGSFNGTHFEAFDNDQSRVVDGCKDYA 286  
Qy 300 AVTPEIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRL 349  
Db 287 LOTFTNTDPTYGSALGIAWASNWEISA-FVPTNP---WRSSMSLVKRFSL 332

## RESULT 6

US-09-331-581-23  
; Sequence 23, Application US/09331581  
; Patent No. 6130070  
; GENERAL INFORMATION:  
; APPLICANT: TOHDA, Hideki  
; APPLICANT: HAMA, Yuko  
; APPLICANT: KUMAGAI, Hiromichi  
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE  
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS  
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF  
; FILE REFERENCE: 0059-1142-OPCT  
; CURRENT APPLICATION NUMBER: US/09/331,581  
; CURRENT FILING DATE: 1999-06-30  
; EARLIER APPLICATION NUMBER: PCT/JP98/04929  
; EARLIER FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: JP 9-314608  
; EARLIER FILING DATE: 1997-10-31  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 332  
; TYPE: PRT  
; ORGANISM: Schwanniomycetes occidentalis  
US-09-331-581-23

Query Match 10.2%; Score 291; DB 4; Length 332;

Best Local Similarity 28.1%; Pred. No. 2e-18;  
Matches 94; Conservative 58; Mismatches 125; Indels 58; Gaps 17;

QY 40 RAVYHMTTPSGWLCDDPQRPV--TTHGAYQLYLYHSDQNG---PGGWDHASTTGDVAFTH 94  
DB 26 RPLHETPEKGMNDPNSGTFYDRTAKTHLYFOYNPNATAWGQPLYWGHATSNDLVHWE 85  
QY 95 HG-TVMPLRDFPVGSGSAVGTANTAGCAGA-----VVALATQPTDGVKRYQOYLY 147  
DB 86 HENAIGPEHDNEIGFSGSIVVDHNTSGFNSIDPNQRIAYI---TNNPDLQTDIA 142  
QY 148 WSTDGGFTTALPDPVIVNTDGRAATPAEIAENAEWFRPKIHW-----DTARGEWC 200  
DB 143 FSLDGGYTFKYNPNVIDV-----SSNQFRPKVFWHERFKSMDHCSE--- 187  
QY 201 VIGRLRYAIFYSPNLRDWTLRNFD--YPNHALGIECPDLFEITADGCTR-HWVLAAS 257  
DB 188 -TARVKIQIF-GSANLKNVLSNFSGGYGNQYG---MSRLIEVPFENSCKSKWYMF 242  
QY 258 MDAYGTGLPM---TYAYMTGTWDEGFHADDLTPOWLDGWDWYAAVTPSIDAPETKRL 314  
DB 243 INP---GSLPGGSINQYFVGDFDGFQFVDDSDQTRFVDIGKDFYAFQTFSEV---EHGVL 296  
QY 315 ATAMNNRYAARDVPTDASDGYNGONSIVRELRL 349  
DB 297 GLAWSNQYADQ-VPTNP---WRSSTSLARNYTL 327

## RESULT 7

US-08-245-809-1  
; Sequence 1, Application US/08245809  
; Patent No. 5665579  
; GENERAL INFORMATION:

; APPLICANT: Fitzmaurice, Leona  
; APPLICANT: Mirkov, Erik  
; APPLICANT: Butler, William O.  
; APPLICANT: Dickinson, Craig D.  
; APPLICANT: Elliott, Kathryn J.  
; APPLICANT: Konno, Yoshihiro  
; TITLE OF INVENTION: NOVEL INVERTASE GENE(S) AND USES THEREOF  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Fitch, Even, Tabin & Flannery  
; STREET: 135 South LaSalle Street, Suite 900  
; CITY: Chicago  
; STATE: IL

; COUNTRY: U.S.A.  
; ZIP: .60603-4277  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/245,809  
; FILING DATE: 17-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/771,331  
; FILING DATE: 03-OCT-1991  
; APPLICATION NUMBER: US 07/660,344  
; FILING DATE: 22-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 51514  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-552-1311  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-245-809-1

Query Match 9.7%; Score 275; DB 1; Length 635;

Best Local Similarity 24.2%; Pred. No. 1.5e-16;  
Matches 137; Conservative 76; Mismatches 236; Indels 118; Gaps 28;

QY 2 TPAISRAVYLAQAGACALALIFGAVPPAARASAPGS-LRAVYHMTTPSGWLCDDPQRPVT 60  
DB 65 STAPPSRGVSQGVSDKTFRDVAGASHVYANSLMSWORTAYHFQPKNNMNDPNGLY 124  
QY 61 THGAYQLYLYHSDQNN--GPGWDHASTTGDVAFTHGTVMP--LRPD-----FPWWSG 111  
DB 125 HKGWYHLFYQYNPDSAIWNTWGHAVSKDLI---HWLYLPFAMVPDQWYDINGVMTGS 180  
QY 112 AVVGTANTAGFAGAVVALATQPTDGVKRYQOYLYWSTDGGFTTALDPPVI--VNTDG 169  
DB 181 ATI-----LPDQQLMMLYTGTDYD--VQVQNLAIPAN-----LSDPLLDWVKFKG 224  
QY 170 RAATTPAEIENAEWFRPKIHW--DTARGEWCV---IGRLRYAIFYSPNLRDWTLRN 224  
DB 225 NPVLVPPPGIGVKDFRDPITANTGPONGWLLTIGSKIKGTGVALVYETSNFTSKL--- 281  
QY 225 FDYPNHALGGI---ECPDLFEITA-----DDGTRHWVLAASM-----DAYGIGL 265  
DB 282 LDGVLHAVPCTGMWECVDFYVSTKKTNGLDTSYNGPGVKH-VLKASLDDNKQDHVAIG- 339  
QY 266 PMTYAYWTGTWDEGFHADDLTPOWLDGWDWYAAVTPSIDAPETKRLAIAAMNNKYA 325  
DB 340 --TYDLGKNKWTDPNPELDCGIGLRIDYG-KYASKTF--YDPKKERRVLMGWIGETDSE 394  
QY 326 ARDVPPTDASDGYNGONSIVRELRLARPGGWYTLTSPVAALTNVYVTTATTLTLDRTVD-- 383  
DB 395 SADL-----QKGNASVQSIPRTVLYDKKTG--THLLOWPVEETESLRVGDPTV--KQVDLQ 446  
QY 384 -GSVLPWNGRAYEIELDIAWDT-----ATNVGISVSGRSPDGRHTNIGKYG--- 429  
DB 447 PGSIELRLVDSAAELDIEASFVQKVALQIIEADHVGFCSTSGCAASRGILGPGFVIV 506  
QY 430 -AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPCARSV 463  
DB 507 IADQTLSELTPYFYISKAGADGRAETHFCADQTRSEAFPGVKQYGVSSVPLDGEKHS- 565  
QY 464 HLRILVDTSQVEFVNAGHTVLSQOVH 490  
DB 566 -MRLLDVHSIVESFAOQGRVTITSRIY 591

RESULT 8  
US-08-296-624-2  
; Sequence 2, Application US/08296624  
; Patent No. 5658773  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Klann, Ellen  
; TITLE OF INVENTION: Tomato Acid Invertase Gene  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Weber  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296.624  
; FILING DATE: 26-AUG-1994  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/770,970  
; FILING DATE: 07-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A.  
; REGISTRATION NUMBER: 31,677  
; REFERENCE/DOCKET NUMBER: 2307E-036510US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-543-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 636 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-296-624-2

Query Match 9.6%; Score 274.5; DB 1; Length 636;  
Best Local Similarity 24.1%; Pred. No. 1.6e-16;  
Matches 137; Conservative 76; Mismatches 236; Indels 119; Gaps 28;  
Qy 2 TPAISRRVLOGAGAGALALIFGGAVPPAARASAFGS-LRAVYHMTTPSGWLCDPQRPVT 60  
Db 65 SPAPSRGVSQGVSDKTPRDVAGASHVSYANSNAMLSQRTAYHFQPKQKNMNDPNGPLY 124  
Qy 61 THGAYQLYLHSDQNN--GPGGWDHASTDGVAFTHHGTVMV--LRPD-----FPVWSGS 111  
Db 125 HKGWTHLEFYQYNPDSAINWNTTGHVASKDLI---HWLYLFFAMVPQWYDINGVWTGS 180  
Qy 112 AVVGTTANTAGFGAGAVVALATQPTDGVKRYOEXLYLWSTDGFTTALPDPIV---VNTD 168  
Db 181 ATI-----LPDQGMILYTGDDTDY--VOVONLAYPAN-----LSDPLLLLDWVKEK 224  
Qy 169 GRAATTPTAEIENAEWFRPKIHW-DTARGEWCV-----IGRLRYAAFTSPNLRDWTLLRR 223  
Db 225 GNPVLVPPPGVGRDPRDPTTAWTGPQNGWLLTIGSKIGTGVALVYTSNFTSKL-- 282  
Qy 224 NFDPNHALGII---ECDFLEITA-----DDGTRHWLAASM-----DAYGIG 264  
Db 283 -LDGVLHVAVPGTGMWECVDFYFPVTKTNGLDTSTNGFGVKH-VLKASLDNDKQDHYAIG 340  
Qy 265 LPMTYAVWTGWDGQFHADDLTPQWLDGWDWVAATWPSIDAPETKRLAIAMNNKY 324  
Db 341 ---TYDLGKNKWTDPNPELDCGIGLRLDYG-KYYASKTF--YDPRKRRVLWGWTGETDS 394

Qy 325 AARDVPTDASDGYNGQNSIVRELRLARQPGWYWTLLSTPVAALTNYVTATTLTPORTVD- 383  
Db 395 ESADL---QKGWASVQSIPRTVLYDKKTG---THLQWPVEEIESLRVGDPTV--KQVDL 446  
Qy 384 --GSVLPWNGRAYEIELDIAWDT-----ATNVGISVGRSPDGTGRHTNICKYG-- 429  
Db 447 QPGSTELLRVDSAAELDIEASEFVDKVALQGIIEADHVGFSCSTGGGAASRGILPGFVI 506  
Qy 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPGARS 462  
Db 507 VIADQTLSELTPVYFYISKGADGRAETHFCADQTRSSEAPGVGKQVYGVSSVPVLGEEKS 566  
Qy 463 VHLRILVDTQSVVEFVNAGHTVLSQQVH 490  
Db 567 --MRLLVDSIVESFAQGRRTVITSRIY 592  
RESULT 9  
US-09-134-001C-3895  
; Sequence 3895, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3895  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-3895

Query Match 8.3%; Score 235.5; DB 4; Length 492;  
Best Local Similarity 23.5%; Pred. No. 4e-13;  
Matches 123; Conservative 69; Mismatches 202; Indels 129; Gaps 24;  
Qy 40 RAVYHMTTPSGWLCDPQRPVTHGAYQLYLHS-----DONNGPGGWDHASTDGVAFTHH 95  
Db 35 RQTFHIQPEMGLLNDPENGLIYNGHY--YISHQWFLPGLAVHGLKYWFNYKSKDLLHFEQ 92  
Qy 96 GTVMPLRPD-----FPVWSGSVAVGTANTAGFGAGAVVALATQPTDGVKRYOEQ--YLYW 148  
Db 93 GTL--LKPDTKYDSHGVYSGSAF-----EYQNHLYMYT 124  
Qy 149 STDGGFTTALPDPIV--VNTDGRAATTPAEIENAE-----WERDPKTHWDTARGEWC 200  
Db 125 GNRDQHNRRISSQMIARNKDKTEKPKVHQQEYTSKSHFRDPKFV--EKNLSQLYA 182  
Qy 201 VIG-----RLRYAFTYSPNLRDWTLLRRFDYPNHALGII-----ECDFLEIT 244  
Db 183 ILGAQNEENEMGLLYRSQDVVDW-----HFEGEIKTNLTQFGYMWCEPDYFRLS 232  
Qy 245 ADDGTRHWVLAASMDAYGIGLPMY--AYWCTWDGEGFHADDLTPQWLDGWDWYAAVT 302  
Db 233 NKDVT--LMCPQGVAEQDKFRNIYQSGYMGIDLNFNLFHDSHESFQELDNGDFYAPQT 290  
Qy 303 WPSIDAPETKRLAIAMNNWKYAADVPTDASDGYNGQNSIVRELRLARQPGWYWTLLST 362  
Db 291 F--VDA-DGQRLIGWGMG---LPDTEYPTD-KEGWAHCLTIPRVLTENG-----KLQR 338  
Qy 363 PYAAL-----TNVTATTLTPDRTVDGSA-----VLPWNGRAYEIELDTATNNGVS 412  
Db 339 PKQLEDLRTNKETAL-----GYANKFKRLHPYEGKQYEMIIDILENDASEIYFE 389  
Qy 413 VGRSPDGTGRHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSV-----HL 465

Db 390 LRSSRSESLITYNKHNKLTLDRTDS-----GTLPSNVDTGTRSTRILDSPKLQ 439  
QY 466 RILVDTSQSEVFNAGHTVLSQOVHFAEGDTGISLYTDGGPAH 508  
Db 440 QIPVDTSIEIFCNDGERVLTSRIFPNEDATGKIKASTESQVY 482

## RESULT 10

US-08-860-091A-2

; Sequence 2, Application US/08860091A

; Patent No. 6057494

; GENERAL INFORMATION:

; APPLICANT: Koops, Dr. Andries Jurriaan

; APPLICANT: Van der meer, Dr. Ingrid Maria

; APPLICANT: Van Tunen, Dr. Arjen Johannes

; TITLE OF INVENTION: DNA sequences encoding carbohydrate

; TITLE OF INVENTION: polymer synthesizing enzymes and method for producing transgen

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LADAS &amp; PARRY

; STREET: 26 WEST 61 STREET

; CITY: NEW YORK

; STATE: NY

; ZIP: 10023

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/4" Floppy disk 1.44MB

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Microsoft Windows for Workgroups

; SOFTWARE: WordPerfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/860,091A

; FILING DATE: 02-JULY-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/NL/00012

; FILING DATE: 08-JANUARY-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Cord, Janet I.

; REGISTRATION NUMBER: 33,778

; REFERENCE/DOCKET NUMBER: U-011329-3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 708-1935

; TELEFAX: (212) 246-8959

; TELEX: 233288

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 630 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-860-091A-2

Query Match 7.2%; Score 205.5; DB 3; Length 630;  
Best Local Similarity 21.2%; Pred. No. 3.1e-10;  
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps 29;

QY 20 ALIFGGAVPPAARASAPGSL-----RAVYHMTPPSGWLCDDPQRPVTHGAYQLYYL 70  
Db 65 ATTFVTQLPNIDLKRVPGKLDSSAEVEWQSTYHFQPKNFISDPDGPVHMGWYHLFYQ 124  
QY 71 HSDONN--GPGGDHASTTGDVAFTHHGTVMPP--LRPDPFVWSSGSAVGTANTAGFCAGA 126  
Db 125 YNPQSAIWNITWGHVSVDKMINWFH----LPFAMVPDH--WYDIEGVMTGSATVLPNGQ 178  
QY 127 VVALATQPTDGVKRYEQOYLWSTDGGFTTALPDPIV---NTDGRAATTAEIENAEW 183  
Db 179 IIMLYS--GNAYDLSOVQCLAYAVNSS-----DPLLEWKYEGNPVLLPPPGVGKYD 229  
QY 184 FRDPKIHMDTARGEWCVCIG-----RLRYAAFTYSPNLRDWTLRNFDYPNHALGGIECP 238  
Db 230 FRDPSTLWSPDGEYRWVNGSKHNETIGCALIYHTTNTFHFELKEEVLHAPHTGMWECV 289  
QY 239 DLFEIT-----ADDGTRH--WVLAASMDA-----YG 262  
Db 290 DLYPSTVHTNGLDVNDGNPNVKYVLKQSGDEDRHDWYAIGSYDIWNKWPDPDPENDVG 349  
QY 263 IGLPMTYAYWTGTWDCGEQFHAD-----DLTPQWLDGWDWYAAVTWP 304  
Db 350 IGL-----RYDGFKEYASKTFYDQHKRRVLWGYVGETDPQKYDLSKGANILNIP 400

Qy 305 S---IDAPETKRLAIAM-----MNNWKY-AARDVPTDASDGYNGONSIVRELRAROP 353  
Db 401 RTVVLDD-ETKTLNLPDIEETENLRSKKYDEKDV-----ELR-----P 439  
Qy 354 GWTYLLSTPVAALTYN-VTATTLTLPDRTVDGSVLPWNGRAYEIELDIANDTATNNGIS 412  
Db 440 G---ALVPLETGATQLDIVATFEIDQKMLEST-----LEADVLENCITTSSEG-S 484  
Qy 413 VGRSPDGTNRHNTNIGRYGADLVXDRGPS-----DLAG-----YSLAPYSR----- 451  
Db 485 VARSV-----LGPFGVVLADAQRSEQLPVYFVIANDIDGTSRTYFCADETRSSKDV 537  
Qy 452 -----AAAPIDGARSVHLRLVDTOSVEFVNAGHTVLSQQVH 490  
Db 538 VGVWYGVSSVPVLPGEK-YNMRLLVHDHSIVEGFAQNGRTVVTSRVY 582

## RESULT 12

US-08-860-091A-4  
; Sequence 4, Application US/08860091A  
; Patent No. 6057494

## GENERAL INFORMATION:

APPLICANT: Koops, Dr. Andries Jurriaan  
APPLICANT: Van der meer, Dr. Ingrid Maria  
APPLICANT: Van Tunen, Dr. Arjen Johannus

TITLE OF INVENTION: DNA sequences encoding carbohydrate

TITLE OF INVENTION: polymer synthesizing enzymes and method for producing transgen

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: LADAS & PARRY

STREET: 26 WEST 61 STREET

CITY: NEW YORK

STATE: NY

ZIP: 10023

COUNTRY: USA

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/4" Floppy disk 1.44MB

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Microsoft Windows for Workgroups

SOFTWARE: WordPerfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/860,091A

FILING DATE: 02-JULY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/NL/00012

FILING DATE: 08-JANUARY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Cord, Janet I.

REGISTRATION NUMBER: 33,778

REFERENCE/DOCKET NUMBER: 0-011329-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 708-1935

TELEFAX: (212) 246-8959

TELEX: 233288

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 615 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-860-091A-4

Query Match 6.6%; Score 189; DB 3; Length 615;

Best Local Similarity 20.0%; Pred. No. 9.4e-09;

Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHHAGYQLYLHSDQNN--GPGGWDHASTDGVAFTHGTT 97  
Db 88 RTAFHFQPAKNEIYDPDQQLFHMGMWYHMFYQYNPYAPVWGNMNSGHSVKDMINWYE--- 144  
Qy 98 VMLPRDPFPVWSSGSAVGTANTAGGAGAVVALATQPTDGVKRYQEQYLYWSTDGGFTFT 157  
Db 145 -LPVAVPTEWYDIEGVLGGSTTVLPNGQIFALYTGNAFDSQLO-----CK 190

Qy 158 ALP-----DPVIVN-----TDGRAATTPAETENAEWERDPKIHWDTARGEVWCVIGRLR--- 206  
Db 191 AVPNVNSDPLLEWVKYEDNPILYTPPGI-GLKDYRDPSTVWTGPDGKHRMIMGTKRGNT 249  
Qy 207 --YAAFYTSPNLRDWTLRNFDYPNHALGGI---ECPDLFETADDGTRHWWVLAASMDAY 261  
Db 250 GMLVYYT---DYNTYELLEDPLHSVPNTDMWECVDFYPVSLTND---ALDMAA 300  
Qy 262 GIGLPMYAYWGTGWDGEQFHADDLTPQWLDGWDWYAAVWTWPSI-----DAPE----- 310  
Db 301 GSGIKHVI---KESWEGH-----GMDWYSIGTYDAINDKWTDPNPELDVGI 343  
Qy 311 -----TKRLAIAMNNWKYAAARDVPTDASDGYNGONSIVRELRARQ 352  
Db 344 GLRCDYGRFFASKSLYDPLKKRRITW-----GYVGESDSA-DQDLRS- 384  
Qy 353 PGGWYLLSTPVAALTNVVTATTLTLPDRTVDGSVAVLPW-----NGRAYE----- 396  
Db 385 --GWATV-----YNVGRTIVLDRKT-GTHLLHMPVEEVESLRYNGOEFKEIKLEP 431  
Qy 397 ---IELDIAMDATNVGI-----SVGRSPDGTNRHNTNIGRY 428  
Db 432 GSIILPDIG--TATQIDIVATFEVDQAAALNATSETDDIYGCTTSLGAAQRG---SLGPF 485  
Qy 429 GADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSYH-----464  
Db 486 GLAVLADGTLSELT---PVYFYIAKKADGGV-STHFCTDKLRSSLDYDGERVYGGTVP 540  
Qy 465 -----LRLIVDTQSVEFVNAGHTVLSQQVH 490  
Db 541 VLDDEELTMLLLVDHSIVEGFAQGGRTVITSRAY 574

## RESULT 13

US-09-245-323A-8

; Sequence 8, Application US/09245323A

; Patent No. 6365800

GENERAL INFORMATION:

APPLICANT: Caimi, Perry G.

TITLE OF INVENTION: Transgenic Crops Accumulating Fructose Polymers and  
TITLE OF INVENTION: Methods for Their Production

FILE REFERENCE: BBI082 US NA

CURRENT APPLICATION NUMBER: US/09/245,323A

CURRENT FILING DATE: 1999-02-05

PRIOR APPLICATION NUMBER: 60/077,727

PRIOR FILING DATE: 1998-03-12

NUMBER OF SEQ ID NOS: 8

SOFTWARE: MICROSOFT OFFICE 97

SEQ ID NO 8

LENGTH: 615

TYPE: PRT

ORGANISM: Helianthus tuberosus

US-09-245-323A-8

Query Match 6.6%; Score 189; DB 4; Length 615;

Best Local Similarity 20.0%; Pred. No. 9.4e-09;

Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHHAGYQLYLHSDQNN--GPGGWDHASTDGVAFTHGTT 97  
Db 88 RTAFHFQPAKNEIYDPDQQLFHMGMWYHMFYQYNPYAPVWGNMNSGHSVKDMINWYE--- 144  
Qy 98 VMLPRDPFPVWSSGSAVGTANTAGGAGAVVALATQPTDGVKRYQEQYLYWSTDGGFTFT 157  
Db 145 -LPVAVPTEWYDIEGVLGGSTTVLPNGQIFALYTGNAFDSQLO-----CK 190  
Qy 158 ALP-----DPVIVN-----TDGRAATTPAETENAEWERDPKIHWDTARGEVWCVIGRLR--- 206  
Db 191 AVPNVNSDPLLEWVKYEDNPILYTPPGI-GLKDYRDPSTVWTGPDGKHRMIMGTKRGNT 249  
Qy 207 --YAAFYTSPNLRDWTLRNFDYPNHALGGI---ECPDLFETADDGTRHWWVLAASMDAY 261

Db 250 GMLVLYVYTT-----DYTNVELLDPLHSPVNTDMWECVDFYVNSLTNDS-----ALDMAAY 300  
QY 262 GIGLPMYAYWTGDEGFHADDLTPOWLDGWDWYAAVTPSI-----DAPE----- 310  
Db 301 GSGIKHVI-----KESWEGH-----GMDWYSIGTYDAINDKWTPDNPELDVGI 343  
QY 311 -----TKRLAIAMNNWKYAAARDVPTDASGYNGQNSIVRELRLAQ 352  
Db 344 GLRCDYGRFFASKSLYDPLKKRRTW-----GYVGESDSA-DQDLRS- 384  
QY 353 PGGWYLLSTPVAALNYVTATTLPTDRTVDSAVLPW-----NCRAYE----- 396  
Db 385 -GWATV-----YNGRTIVLDKRT-GPHLLHWPVEEVESRYNGQEKETKLEP 431  
QY 397 ---IELDIAMDTATNVI-----SVGRSPDGTGRHTNIGKY 428  
Db 432 GSIIPLDIG-TATQLDIVATFEVQOALNATSETDDIYGCTISLGAORG-----SLGPF 485  
QY 429 GADLYVDRGPSLAGYSLAPYSRAAIPDPGARSVH----- 464  
Db 486 GLAVLADGTLSELT-----PVYFIARKADGV-STHFCTDKLRSLDYDGERVYVGGTVP 540  
QY 465 -----LRLIVDTOSVEVFNAGHTVLSQOVH 490  
Db 541 VLDDEELTMRLLVDHSIVEGFAQGGRTVITSRAY 574

## RESULT 14

US-09-019-385-2  
; Sequence 2, Application US/09019385  
; Patent No. 6147280  
; GENERAL INFORMATION:  
; APPLICANT: Smeekens, Josephus C.M.  
; APPLICANT: Ebskamp, Michael J.M.  
; APPLICANT: Geerts, Hendrikus A.M.  
; APPLICANT: Weisbeek, Petrus J.  
; TITLE OF INVENTION: Production of Oligosaccharides in  
; TITLE OF INVENTION: Transgenic Plants  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen O'Connor Johnson & KindnessPLLC  
; STREET: 1420 Fifth Avenue, Suite 2800  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/019,385  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/479,470  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: NL 1000064  
; FILING DATE: 05-APR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brogerick, Thomas F.  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: ARNO18589  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-224-0709  
; TELEFAX: 206-224-0779  
; TELEX: 4938023  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 626 amino acids  
; TYPE: amino acid

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-019-385-2  
Query Match 6.0%; Score 169.5; DB 4; Length 626;  
Best Local Similarity 21.8%; Pred. No. 5.7e-07;  
Matches 136; Conservative 76; Mismatches 235; Indels 177; Gaps 36;  
QY 16 AGALALIFGAV-----PPAARASAPGS-----LRAVYHMTPPSGWLCDPQR 57  
Db 40 ASAMAVVVVGATLLAGLRMEQAVDEEAAGGFPWSNEMLOWQSGYHFQAKNYMSDPNG 99  
QY 58 PVTTHGAYOLYLHSDQNNPG-GWDHASTDGVATTH-----HGTWMLRPDPFVW- 108  
Db 100 LMYRYGWYHMFY---QYNPVGTDWD---DGMGHAHVSRNLVQWRTLPIDAMADQWY 150  
QY 109 -----SGSAVV---CTA---NTAGFGAGV---VALATOPTDGV---RKQEOYLYW 148  
Db 151 DILGVLSGSMVLPNGTVIMITYGATNASAVEQCITATPADPNDPLLRKWKHPANPVIW 210  
QY 149 STDGGFTTALPDPIVINTDGRAATTPAETENAEWFRDPKIH-W-DTARGEWCVIGRL-- 205  
Db 211 SPPG-----VGTKD-----FRDPTAWYDESDETWRLLGSKDD 244  
QY 206 -----RYAFTYSPNLRDWTLRNFDYPNHALGGTECPDLFEI---TADDGTR--HWVL 254  
Db 245 HDGHDGCIAMMYKTKDFLNVELIPGILHRVVRTGEWECIDFYVGRSSDSSSEMLH-VL 303  
QY 255 AASM-----DAYGIGLPMYAYWTGWDGEQPHADDLTPOWLDWG---WDW---YAAVT 302  
Db 304 KASMDDERHDYSLG---YDASAANTW-----TPIDPE-LDLGIGLRYDNGKFTASTS 352  
QY 303 WPSIDAPETKRLAIAMNNWKYAADVPTDASGYNGQNSIVRELRLARQPGGWYTLST 362  
Db 353 F--YDPAKNRRVLMGVGEVDSKADV---VKGNASIQSPRTVALDEKTR--TNLLW 404  
QY 363 PVAALTNVYTTTLPDRVD-GSAV-LPWNGRAY-----EIELDIAMDATN---VGI 411  
Db 405 PVEEIETRLNATELTDTVTINTGSIHIPLRQGTTHARHAEASFHLDASAVAALNEADV 464  
QY 412 SVGRSPDGTGRHTNIGYGADLYVDRGPSLAGYSLAPYSRAAIPDPG----- 459  
Db 465 NCSSSGGAVNRGALGPFGL-LVLAAG--DRGEQTAIVFYVSKGLDGLGHTSFCDELRS 521  
QY 460 -----ARSHRLILYDTSQSVFVFNAGHTVLSQOVHFAEG---DTGI 498  
Db 522 SRKADVTKRIVGSTVPVLDGEALSMRVLVDHSIVQFDMGGRITMTSRVYPMESYOEAV 581  
QY 499 SLYTDGGAFTG--IVVREIGQA 520  
Db 582 YLFNNTGASVTAERLVVHEMDSA 605

## RESULT 15

US-09-142-623-11  
; Sequence 11, Application US/09142623  
; Patent No. 6337201  
; GENERAL INFORMATION:  
; APPLICANT: Koji YANAI et al.  
; TITLE OF INVENTION: -FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; TITLE OF INVENTION: ISOLATING -FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING  
; TITLE OF INVENTION: -FRUCTOFURANOSIDASE, AND -FRUCTOFURANOSIDASE VARIANT  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible



OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/142.623  
FILING DATE: September 10, 1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 98-0989\*/LC(WMC)/144  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 565 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: No. 6337201 relevant  
TOPOLOGY: No. 6337201 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Penicillium roqueforti IAM7254  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..565  
IDENTIFICATION METHOD: E  
US-09-142-623-11

Query Match 5.2%; Score 147.5; DB 4; Length 565;  
Best Local Similarity 21.8%; Pred. No. 4.9e-05;  
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;

Qy 26 AVPPAARASAPGSL---RAYVHTPPSGWLCDP---QRPVTHGAYQLYVLHSDQNG 77  
Db 11 SAPPNLSLANASLFTWRPRAHLPLPPSGNIGDPCGHYTDPKT--GLFHVGWLYS---G 64

Qy 78 PGWDHASTTDCGVAFTHTGTVMLPRDPFVWMSGSAVGVGTAN-TAGFGAGAVV--ALATQP 134  
Db 65 1SG---ATTDDL---TYKDLNPD---GAPSIAGGKNDPLSVFGSGVIPSGIDGMP 112

Qy 135 T-----DGVRYQEQLYMWSTDGGFTFTALPDPIVINTDGRAATTPAET 179  
Db 113 TLLYTSVSLPIHWSIPYTRGSETQSLAVSYDGGHNETKL-----NQGPIVTPPFAL 165

Qy 180 NAEWFRDPKIH---DTARGEWCVIG-----RLRYAAYTSPNLRD 218  
Db 166 NVTAFRDPYVFQSPILDKSVNSTGTWYVAISGGVHVGPCQFLYRONDADFQYWEYLQ 225

Qy 219 W---TLRNFDPNHALG---GIECPDLFEITADGTRHWVLAASMDAYGIGLP----- 266  
Db 226 WKPELNTTWKGDWAGWGWFENFVGNVFSNAGYSEDEGEIFITLGAEGSGLPIVPOVS 285

Qy 267 ---MTYAYWTGWDGEQFHADLTPOWLDKMGWDYAAV--TWPSIDAPETKR-----L 314  
Db 286 SIRDMLWVTGNVNDGSVTEKPTMAGV-LDWGSAYAAAGKILPASSQASTKSGAPDREI 344

Qy 315 AIAWNNHKY-AARDVPTDSDGNGQNSIVREL-----RLARQ 352  
Db 345 SYVMLTGLDFEQVGFPF-AQONMTGALLPRELNVRTISNVVDNELSRESLTWSRVARE 403

Qy 353 PGWYTLTSTPVAALNTVVTATT-----LPDRTVDGSAVLPN---GRAVEIELDIAW 403  
Db 404 DSGQIDLETMGISISRETYISALTSGSSPVEGKTLNAGAVPFNTSPSKFVLTAISF 463

Qy 404 DT-ATNVGISVGRSPDGRTHNICKY-----GADLYVDRGSDLAGYSLAPYSRAAPI-- 456  
Db 464 PTSARDSGIQAGFQVLSLSLESTTIYQFSNIEIIVDRSNTSAA-----ARTTAGILS 516

Qy 457 --DPG-----ARSVHLRLVDTQSVFVFN 479  
Db 517 DNEAGRLRFLVLRNGKEQVETLELTIVVDNSVLEVYAN 555

Search completed: May 6, 2003, 12:36:32  
Job time : 23.5375 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:34:52 ; Search time 17.5337 Seconds  
(without alignments)  
2563.947 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845  
Sequence: 1 MTPAIRRAVLQAGAGALA.....TDGGPAHPTGIVREIGQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 499.5 | 17.6        | 943    | 10 | US-09-969-362-5    |
| 2          | 468.5 | 16.5        | 1277   | 10 | US-09-969-362-3    |
| 3          | 461   | 16.2        | 923    | 10 | US-09-969-362-1    |
| 4          | 221   | 7.8         | 635    | 9  | US-10-003-392-14   |
| 5          | 206   | 7.2         | 630    | 9  | US-10-003-392-16   |
| 6          | 205.5 | 7.2         | 630    | 9  | US-10-003-392-18   |
| 7          | 202.5 | 7.1         | 433    | 9  | US-09-738-626-6405 |
| 8          | 189   | 6.6         | 615    | 9  | US-10-003-392-17   |
| 9          | 184   | 6.5         | 608    | 9  | US-10-003-392-2    |
| 10         | 176   | 6.2         | 625    | 9  | US-10-003-392-21   |
| 11         | 160   | 5.6         | 600    | 9  | US-10-003-392-20   |
| 12         | 155   | 5.4         | 609    | 9  | US-10-003-392-4    |
| 13         | 147.5 | 5.2         | 565    | 9  | US-09-990-385-11   |
| 14         | 141.5 | 5.0         | 635    | 9  | US-09-990-385-1    |
| 15         | 140.5 | 4.9         | 574    | 9  | US-09-990-385-13   |
| 16         | 133.5 | 4.7         | 495    | 9  | US-10-003-392-8    |
| 17         | 131.5 | 4.6         | 498    | 9  | US-09-738-626-4737 |
| 18         | 126   | 4.4         | 390    | 9  | US-10-003-392-6    |
| 19         | 126   | 4.4         | 471    | 9  | US-10-003-392-10   |

|    |       |     |       |    |                    |                    |
|----|-------|-----|-------|----|--------------------|--------------------|
| 20 | 117   | 4.1 | 451   | 9  | US-09-712-363-171  | Sequence 171, Appl |
| 21 | 115   | 4.0 | 3739  | 9  | US-09-860-846-33   | Sequence 33, Appl  |
| 22 | 115   | 4.0 | 3739  | 9  | US-09-988-384B-33  | Sequence 33, Appl  |
| 23 | 115   | 4.0 | 3739  | 10 | US-09-861-289-33   | Sequence 33, Appl  |
| 24 | 115   | 4.0 | 11877 | 9  | US-09-860-846-6    | Sequence 6, Appl   |
| 25 | 115   | 4.0 | 11877 | 10 | US-09-861-289-6    | Sequence 6, Appl   |
| 26 | 115   | 4.0 | 12199 | 9  | US-09-988-384B-6   | Sequence 6, Appl   |
| 27 | 113   | 4.0 | 296   | 10 | US-09-960-472-1    | Sequence 1, Appl   |
| 28 | 112   | 3.9 | 666   | 10 | US-09-771-161A-136 | Sequence 136, Appl |
| 29 | 111   | 3.9 | 3472  | 9  | US-10-027-806-4    | Sequence 4, Appl   |
| 30 | 111   | 3.9 | 3472  | 9  | US-10-034-623-4    | Sequence 4, Appl   |
| 31 | 111   | 3.9 | 3472  | 9  | US-10-027-801-4    | Sequence 4, Appl   |
| 32 | 110.5 | 3.9 | 983   | 10 | US-09-771-161A-227 | Sequence 227, Appl |
| 33 | 107   | 3.8 | 682   | 10 | US-09-798-791-2    | Sequence 2, Appl   |
| 34 | 106.5 | 3.7 | 1352  | 9  | US-09-784-554B-2   | Sequence 2, Appl   |
| 35 | 106   | 3.7 | 3519  | 9  | US-09-808-880-4    | Sequence 4, Appl   |
| 36 | 106   | 3.7 | 4613  | 9  | US-09-860-846-31   | Sequence 31, Appl  |
| 37 | 106   | 3.7 | 4613  | 9  | US-09-988-384B-31  | Sequence 31, Appl  |
| 38 | 106   | 3.7 | 4613  | 10 | US-09-861-289-31   | Sequence 31, Appl  |
| 39 | 105.5 | 3.7 | 811   | 9  | US-10-011-366-7    | Sequence 7, Appl   |
| 40 | 105.5 | 3.7 | 2710  | 9  | US-10-011-366-6    | Sequence 6, Appl   |
| 41 | 105   | 3.7 | 742   | 9  | US-10-077-111-11   | Sequence 11, Appl  |
| 42 | 105   | 3.7 | 1344  | 9  | US-09-738-626-6888 | Sequence 6888, Ap  |
| 43 | 104   | 3.7 | 655   | 9  | US-10-124-880-14   | Sequence 14, Appl  |
| 44 | 103.5 | 3.6 | 559   | 10 | US-09-725-735A-15  | Sequence 15, Appl  |
| 45 | 103.5 | 3.6 | 630   | 10 | US-09-725-735A-16  | Sequence 16, Appl  |

ALIGNMENTS

RESULT 1  
US-09-969-362-5  
; Sequence 5, Application US/09969362  
; Patent No. US20020076790A1  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rabbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6, b-D-Fructan Hydrolase Enzyme And  
; FILE OF INVENTION: Process For Using The Enzyme  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/969,362  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: 09/397,885  
; PRIOR FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: PA 1998 01623  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: 60/101,615  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/111,675  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-09-969-362-5

|                       |       |  |       |            |     |        |     |
|-----------------------|-------|--|-------|------------|-----|--------|-----|
| Query Match           | 17.6% | Score  | 499.5 | DB         | 10  | Length | 943 |
| Best Local Similarity | 30.8% | Pred. No.  | 7e-33 |            |     |        |     |
| Matches               | 157   | Conservative   | 67    | Mismatches | 223 | Indels | 63  |
| Gaps                  | 19    |  |       |            |     |        |     |
| QY                    | 40    | RAVYHMTPPSGWLCDDPQRPVTHGAYQLYLYHSDQNGPGWDHASTTDCVAFTHHGTVM | 99    |            |     |        |     |
| Db                    | 383   | RPDYHYPARGASDSDNGLVYVEGHLF--HQD---GGTWAHAVSTDLVHKRLPIAL    | 436   |            |     |        |     |
| QY                    | 100   | PLRPDPVWVGSGAVGVCNTAGF-----GAGAVVALATQTDGVRYQEQYLYWSTDGDF  | 154   |            |     |        |     |
| Db                    | 437   | PWDLGHVWVGSAVADLHNASGLFADSGKGLIAYTTSYNPDRPNGNQRIGLAYSKDGR  | 496   |            |     |        |     |



Db 446 PNDHGHVWSGSAVADMTNASLFGDSGKGLIAYTTFNPDSPNGNORIGLAYSKDOOR 505  
Qy 155 TFTAALPD-PVIVNTDGRAATPAETENAEW-PROPKIHWDTARGEMVVCVIGRLRAAEVT 212  
Db 506 TWEYSKERPIVENPKSGN-----EAGNDFRDPKVIIRDENNRMVWVSGGDHIREYT 560  
Qy 213 SPNLRDWTLRNFDYPNHALGGI-BCPDLEITADDGT--RHWVLAASMDAYGIGLPMW- 268  
Db 561 STNLLDWTLDNWGYDGVYRGVWECPLDLPV-DGT SQKKVMMISTGAN---PKTG 615  
Qy 269 ---YAYWTG--TWGEQPHADDLTPQWL--DWGWDWYAAVTPSIDAPETKRLAIAMNN 321  
Db 616 GSDAEYFYGHLTADG-KFVNNDNPAKVLRTDFGKFEYASMSFANN--PDHRTVMAMWNTN 672  
Qy 322 WKYAADVPDASDGYNGNSIVRELRLARQPGWYTLISLTPVAALTNVYATTTLPDPT 381  
Db 673 WDYFPA-FPT---SNWKELIPREVSLLVTTEDG-IRWVQPIKEL-----ESLRKPLYS 722  
Qy 382 VDGSAVLPNRGR-----AYEELDI---AWDTATNVGIVSGRSPDGRHNTNIGKYGA 430  
Db 723 ASKNSVSPSSNLLKGIISGAYEIEAEIETSTVTEFGENIREG--ANQKTVVGYKAS 780  
Qy 431 D--LYVDRGPDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSVVEFVNAGHTVLSQO 488  
Db 781 DSRMFVDRTAGSETDFSNLFSKKHEAPTOMENNRIMRLVDESSVEAFGNDGKYVFSV 840  
Qy 489 VFAEGDGTGISLYTDGGAHFTGIWVREI 517  
Db 841 IFPDPAASRAMSFYKGGVNNVSLKVHQL 869

## RESULT 4

US-10-003-392-14  
; Sequence 14, Application US/10003392  
; Patent No. US20020170086A1

## GENERAL INFORMATION:

; APPLICANT: Allen, Stephen M.  
; APPLICANT: Calmi, Perry G.  
; APPLICANT: Stoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BBI463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003,392  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 14  
; LENGTH: 635  
; TYPE: PRT  
; ORGANISM: Parthenium argentatum Grey  
US-10-003-392-14

Query Match 7.8%; Score 221; DB 9; Length 635;  
Best Local Similarity 21.5%; Pred. No. 4.5e-10;

Matches 126; Conservative. 82; Mismatches 207; Indels 170; Gaps 29;

Qy 12 QCAGAGALALIFGAGVPAARASAPGSL-----RAYVHMTPPSGWCLDQRPVTH 62  
Db 69 QGA-----ATFTTQLPKIDMKRVPGLDSDGADVQWQORSAYHFQDKNYSIDPDGPMYHM 123  
Qy 63 GAYQLYLHSDQNN--GPGGDHASTTDCGVAFTHTGVTMP--LRPDPFVWSGSAVVGATN 118  
Db 124 GWYHLFYQNPESAIWNTGHSVSKDMINWFH---LPFAMVPDH--WYDIEGVMTGS 177  
Qy 119 TAGFGAGAVVALATOPTDGVKRYOEYLYWSTDGGETFTALPDPIV---NTDGRAATTP 175  
Db 178 ATVLNPGEIFIMLYT--GNAYDLSQVCLAYAVNSS-----DPLLEWKKYEGNPVLLP 228  
Qy 176 AEIENAEWFRDPKIHWDTARGEMVVCVIG-----RLRYAAFTSPNLRDWTLRNFDYPNH 230

Db 229 PPGVGYKDFRDPSTLWLGPDGEYRMVMSKHNETIGCALIYHTTTFELNEEVLHAVP 288  
Qy 231 ALGGIECPDLPEIT-----ADGTR-HWVLAASMDA-----YGICLPWYAYWTCT 275  
Db 289 HTGMEWECVDLYPVSHTHTNGLDMDVNGPNKYVVLKOSGDEDRHDWYAG-----SYD 340  
Qy 276 WDGEQFHADDLTPQ-----WLDWGWDWYAAVTPSIDAPETKRLAIAMNNWYAAARD 328  
Db 341 WYNDKWDYD--PENDVGIGLRYDYG-KFYASKTF-----YDQHKRRVLM----- 383  
Qy 329 VPTDASDGYNGNSIVRELRLARQPGWYTLISLTPVAALTNVYATTTLP----- 378  
Db 384 -----GYVGEDT-PEKYDLTK--GWANILNRTVTVLDTKTNLIQWPIEETELR 432  
Qy 379 ---DRTVD---GSAVLPNWNGRAVE-----FELDIADWTATNVGIS 412  
Db 433 SKKYDKFVDELVRPGSLPLEIGTATQLDIVATFEVDQMMLLESTLEADVFNCTTSVG-S 491  
Qy 413 VGRSPDGRHNTNIGKYGADLYVDRG-----PSDLAG-----YSLAPYSRAAAPID 457  
Db 492 VGRGV-----LGPFGVVVLADAQRTQLPVYFYIAKDTDGTSTRTYFCADETRSSKQVD 544  
Qy 458 PGA-----RSVHLRLIVDTQSVVEFVNAGHTVLSQOVH 490  
Db 545 VCKWYGVSGVPLPNEKYNMRLLDVHSIVEGFAQNGRTVVTSRVY 589

## RESULT 5

US-10-003-392-16  
; Sequence 16, Application US/10003392  
; Patent No. US20020170086A1

## GENERAL INFORMATION:

; APPLICANT: Allen, Stephen M.  
; APPLICANT: Calmi, Perry G.  
; APPLICANT: Stoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BBI463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003,392  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 630  
; TYPE: PRT  
; ORGANISM: Helianthus sp.  
US-10-003-392-16

Query Match 7.2%; Score 206; DB 9; Length 630;  
Best Local Similarity 21.8%; Pred. No. 7.7e-09;

Matches 125; Conservative 81; Mismatches 204; Indels 164; Gaps 29;

Qy 28 PPAARASA-----PGSL-----RAYVHMTPPSGWCLDQRPVTHGAYQL 67  
Db 62 PPAATTFATQLPNIDLRKVPGLDSSAEVWORSAYHFQDKNFISDDPGPMYHGWYHL 121  
Qy 68 YLHSDQNN--GPGGDHASTTDCGVAFTHTGVTMP--LRPDPFVWSGSAVVGATAGFG 123  
Db 122 FYQYNPESAIWNTGHSVSKDMINWFH---LPFAMVPDH--WYDIEGVMTGSATVLP 175  
Qy 124 AGAVVALATOPTDGVKRYOEYLYWSTDGGFTTALPDPIV---NTDGRAATTPAEIEN 180  
Db 176 NGQITIMLYT--GNAYDLSQVCLAYAVNSS-----DPLLEWKKYEGNPVLP 226  
Qy 181 AEWFRDPKIHWDTARGEMVVCVIG-----RLRYAAFTSPNLRDWTLRNFDYPNHALGGI 235  
Db 227 YKDFRDPSTLWLGPDGEYRMVMSKHNETIGCALIYHTTTFELKEEVLHAVPHTGMW 286  
Qy 236 ECPDLFEIT-----ADGTR-HWVLAASMDA-----YGICLPWYAYWTGWDGEQ 280

Db 287 ECVDLPVSTVHTNGLDWDVNGPNVYVLKQSGDEDRHDWYALG---SYDVVNDKWVPPD 343  
QY 281 FHADDLTPOWLWDGWDYAAVTPSIDAPETKRLAIAMNNWKYAAARDVPTDASGYNGQ 340  
Db 344 PENDVGIGLRDFG-KFYASKTF---YDQHKRRRLW-----GYVGE 381  
QY 341 NSVRELRLARQGGWYTLTSTPAAL-----TNY-----VTATTLTLPDRTVD----- 383  
Db 382 TD-PORYDISK---GWANILNIPRTVLDTKTNLIQWP IETENLSRKYDEFKDV 437  
QY 384 -GSAVLPNNGRAYE-----TLDIANDTATNVGTSVGRSPDGTHTN 424  
Db 438 RPSGLPLEIGTATQDIVATFEIDQKMLESTLEADVLFNCITSEG-SVARGA----- 489  
QY 425 IGRYGADLYVDRGSP-----DLAG-----YSLAPYSR-----AAPI 456  
Db 490 LGPGVGVVLADAQRSEQLPVYFIANDIDGTSRTYFCADETRSSKDVSVGKWKYVSSVPV 549  
QY 457 DPGARSVHLRLVDTSQSVFVNAGHTVLSQQVH 490  
Db 550 LPGEK-YNNRLLVDHSIVGFAQNGRTVVTSRVY 582

## RESULT 6

US-10-003-392-18  
; Sequence 18, Application US/10003392  
; Patent No. US20020170086A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Calmi, Perry G.  
; APPLICANT: Stoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BB1463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003,392  
; PRIOR FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 18  
; LENGTH: 630  
; TYPE: PRT  
; ORGANISM: Helianthus tuberosus  
US-10-003-392-18

Query Match 7.2%; Score 205.5; DB 9; Length 630;  
Best Local Similarity 21.2%; Pred. No. 8.4e-09;  
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps 29;

QY 20 ALIFGCAVPPAARASAPGSL-----RAVYHMTPPSGWLCDDPQRPVTHGAYOLYYL 70  
Db 65 ATTFVQLNIDLKRVPGLDSSAEVWORSYHFQDPKFNFTSDPDGPMYHGWYHLFTQ 124  
QY 71 HSDQNN--GPGGMDHASTTGDGVAFTHTGTVMP--LRPDPFVWMSGSAVVGTTANTAGFGAGA 126  
Db 125 YNPQSAIWNITWGHVSXDMINWFH---LPFAMVPDH--WYDIEGVMGTSATVLPNGQ 178  
QY 127 VVALATQPTDGVKRYQEQVLYNSTDGGFTTALPDPIV---NTDGRATTTAEIENAEW 183  
Db 179 IMLYS--GNAYDLSQVCLAVNNS-----DPLLEWKYEGNPVLLPPPGVGKYD 229  
QY 184 FRDPKTHWDTARGEMWCVIG-----RLRYAAFTSNLRDWTLLRRNFDPNHALGIECP 238  
Db 230 FRDPSTLWSPGDEYRWVMSKHNETIGCALLYHTNTFTFELKEVLAHVHTGMWECV 289  
QY 239 DLFEIT-----ADDGTRH-WVLAASMDA-----YG 262  
Db 290 DLYPSTVHTNGLDVNDGPNVYVLKQSGDEDRHDWYAIQSYDIVNDKWVPPDPENDVG 349  
QY 263 IGLPWTYAWTCTWDEGEQPHAD-----DLTPQWLDWGNWDWYAAVTPW 304

Db 350 IGL-----RYDGKFTYASKTFYDQHKRRVRLNGYVGETDQKVDLSKGNILNIP 400  
QY 305 S-----IDAPETKRLAIAM-----MNNWKY-AARDVPTDASDGYNGONSTVRELRLARQP 353  
Db 401 RTVVLDL-ETKTNLIQWPIEETENLSRKYDEFKDV-----ELR---P 439  
QY 354 GGYWTLTSPVAALTNV-VTATTLTLPDRTVDSAVLPWNGRAYEIELDIAMDATATNVGIS 412  
Db 440 G---ALVPLEIGTATQDIVATFEIDQKMLEST-----LEADVLFNCTTSEG-S 484  
QY 413 VGRSPDGTHTNIGKYGADLYVDRGSP-----DLAG-----YSLAPYSR----- 451  
Db 485 VARSV-----LGPGVGVVLADAQRSEQLPVYFIANDIDGTSRTYFCADETRSSKDV 537  
QY 452 -----AAAPIDPGARSVHLRLVDTSQSVFVNAGHTVLSQQVH 490  
Db 538 VGKWKYVSSVPVLPGEK-YNNRLLVDHSIVGFAQNGRTVVTSRVY 582

## RESULT 7

US-09-738-626-6405  
; Sequence 6405, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 6405  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6405

Query Match 7.1%; Score 202.5; DB 9; Length 433;  
Best Local Similarity 24.8%; Pred. No. 9.1e-09;  
Matches 132; Conservative 49; Mismatches 166; Indels 185; Gaps 32;

QY 38 SLRAVYHMTPPSGWLCDDPQRPVTHGAYQLYLHSDQNNNGP-----GWDHAST--- 86  
Db 11 SLRPAVHTVPPQGRNDPNMGVVDGTLHYYOHD-----PGFPFAPKRTGWAHTTPTL 65  
QY 87 -TDGVAFTHTGTVMP--LRPDPF-----VWMSGSAVVGTTANTAGFGAGAVVALATQPTDGV 138  
Db 66 GQRLQWTH-----LPDALYPDASYDLDDGCGSAGAVFTDGTCLKLYTGNL-----KIDGK 115  
QY 139 RYQEQVLYNSTD-----GSETFTALPDPIVNTDGRAA-TTFAEIAENAEWFRDPKIHWD 192  
Db 116 RR-ATONLVEVEDPTGLMGGIHRSPKNPLI---DGPASGFTP-----HYRDPMLSPD 164  
QY 193 TARGE-WCVIGRLR-----YAAFTSPNLRDWTLLRRNFY-----PNHALGGI- 235  
Db 165 ---GDGKKNVLAQRENLTGAALYRSTDLNWEFSGEITFDLSDAQPGSAPDLVPGYM 221

```

Query Match      6.6%; Score 189; DB 9; Length 615;
Best Local Similarity 20.0%; Pred. No. 1.9e-07;
Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAYVHMTPPSGWLCDPQRPVTHGAYQLYYLHSDQNN--GPGGWDHASTTDCGVAFTHHGT 97
      | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 88 RTAFHFQPAKNFIYDPDQQLFHMGWIMFYQYNPYAPVWGNMSGHSYSKDMINWYE--- 144
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 98 VMPLRPDPFVWMSGAVVGTANTAGFAGAVVALATPQTDGVKRYQEQLYLYNSTDGGFTFT 157
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 145 -LPVAVMPTIEWYDIEGLVSGSTTVLPNGOIFALYTGNANDESOLQ-----CK 190
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 158 ALP-----DPVIVN-----TDGRAATPAEIAENAFWRDPKTHWDTARGEWVCVIGRLR--- 206
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 191 AVPNLSDPILLIEVWKYEDNPILYTPPGI-GLKDYRDPSTVMTGPDGKRRHIMGTGRKNT 249
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 207 --YAAFTSPNLRDWTLRNFDYENHALGGI--ECPDLPFETADDGTRHVVLAASMDAY 261
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 250 GMLVLYITT---DYTNVELLDELPHSVPTDMWECVDYFVPSUTNDS-----ALDMAAY 300
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 262 GIGLPMTYAYWTGWDGEQFHADDLTQWLGDWGDWDAAVTWPST-----DAPE----- 310
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 301 GSGIKHVI---KESWECH-----GMDWYSIGTYDAINDKWTDPNDELVDGI 343
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 311 -----TKRLATAWNNNKYAARDVPTDASDGYNGQNSIVLRRLIARQ 352
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 344 GLRCDYGRFFAASKLYDPLKKRRITW-----GYGESDSA-DODLSR- 384
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Qy 396 EIE-----LDATWDATNVGISVGRSPDGTR---HTN-----IGY 428
||| ||| | : | : || : | :
Db 419 EIEKPGSIIPLDIGSSATQLDIVATFEVDODALKAISETNEEYICTKSWGAAGRGLGPF 478
||| ||| | : | : || : | :
Qy 429 GADLYVDRGSPDLAGY-----SLAPYSR-----AAAPIDPCA 460
||| ||| | : | : ||| ||| | : | :
Db 479 GVAVLAGDTSELTPVYFYIAKNNTDGSVATHCTDKLRSLDYDRERVYGVSTVPVLDG- 537
||| ||| | : | : ||| ||| | : | :
Qy 461 RSVHLRILVDTQSVVEFVNAGHTVLSQQVH 490
||| ||| ||| ||| ||| ||| :
Db 538 EELTMRLLVDSHVVEGFAQGRVTITSRVY 567
||| ||| ||| ||| ||| ||| :

RESULT 10
US-10-003-392-21
; Sequence 21, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Scoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-003-392-21

```

```

Query Match      6.2%; Score 176; DB 9; Length 625;
Best Local Similarity 21.5%; Pred. No. 2.3e-06;
Matches 134; Conservative 77; Mismatches 236; Indels 176; Gaps 34;

QY 16 AGALALIFGGAV-----PPAARASAPGS-----LRAVYHMTPSGMLCDPQR 57
DB 40 ASAMVVVGATLLAGLRMEQAVDEEAAAGFPWSNEMLQORSGYHPQTAKNYMSDPNG 99
QY 58 PVTHGAYQLYYLHSDQNGPG-GWDHASTTDDGVAFTH-----HGTVMPLRDPFPVW- 108
DB 100 LMYRYRGYHMFY-----QYNPVGTDMW-----DGMWGHAVSRNLVQWRTLP IAMVADQWY 150
QY 109 -----SGSAVV-----GTA---NTAGFGAGAV-----VALATQPTDGV-----RKYOEQYLW 148
DB 151 DILGVLSSMTVPLNGIVIMLYTGATNASAVEVOCIATPADPNPDLRRWTKHPANPVIW 210
QY 149 STDGGFTTALPDPIVNTDGRAATTPAEIAENAEWFRDKIHW-DTARGEVWCVIGRL-- 205
DB 211 SPPG-----VGTKD-----FRDPMTAWYDESDETWRTLIGSKDD 244
QY 206 -----RYAAFTSPNLRDWTLRRNFDPYNHALGGIECPDLFEI---TADDGTR-HMWL 254
DB 245 HDGHDGIAMMYTKDFLNYELIFGILHRVVTGEWECIDFPVGRSSDNLSEMLH-VL 303
QY 255 AASM-----DAYGTLPLMTYAYWTGTWDGEGFHADDTLPQWLDWG-----WDW---YAAVT 302
DB 304 KASMDDERHDYISLG---TYDASAANTW-----TPIDPE-LDLIGLIRYDMWGFYASTS 352
QY 303 WPSIDAPETKRLTAWNNKYYAARDVPTDASDCYNGONSITVRELRLAROPGWTYLLST 362
DB 353 F--YDPAKNRRLVMGYGEVDSKRAV-----VKGWASIQSVPRVALDEKTR---TNLLW 404
QY 363 PVAALTNYVTATTPLDPTVDGSAVLPE---WNGRAYEIELDIAWDTA-----TNVGIS 412
DB 405 PVBEIETLRNLATBELTDVNTNGSVIHPILROGTFOLDIEASFHLDASAAVAALEADVGN 464

```

```

Qy 413 VGRSPDGRTHNICKYGADLYVDRPGPSDLAGYSLAPYSRAAAPIDPG-----459
      |      |      |      |      |      |      |      |      |
Db 465 CSSSGGAVNRGALGPPFGL-LVLAAG--DRRGEQTAVFYVSRGLDGLHTSFQDELRRSS 521
Qy 460 -----ARSVHLRIILVDTSVEFVNAGHTVLSSQVHPAEG--DTGIS 499
      |      |      |      |      |      |      |      |      |
Db 522 RAKDVTNRVIGSTVPVLDGEGALSMRVLVDHSIVOGEDMGGRITMTSRVYPMESYQEARVY 581
Qy 500 LYTDGGPAHFTG--IVVREIGQA 520
      |      |      |      |      |      |      |      |
Db 582 LFNNATGASVTAERLVVHEMDSA 604

RESULT 11
*US-10-003-392-20
: Sequence 20, Application US/10003392
: Patent No. US20020170086A1
: GENERAL INFORMATION:
: APPLICANT: Allen, Stephen M.
: APPLICANT: Calimi, Perry G.
: APPLICANT: Stoop, Johan M.
: TITLE OF INVENTION: Fructan Biosynthetic Enzymes
: FILE REFERENCE: BB1463 US NA
: CURRENT APPLICATION NUMBER: US/10/003,392
: CURRENT FILING DATE: 2001-10-30
: PRIOR APPLICATION NUMBER: 60/244,273
: PRIOR FILING DATE: 2000-10-10
: PRIOR APPLICATION NUMBER: 60/269,543
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 20
: LENGTH: 600
: TYPE: PRT
: ORGANISM: Triticum aestivum
*US-10-003-392-20

```

| Query Match           | 5.6%   | Score 160;  | DB 9;           | Length 600; |
|-----------------------|--------|---|-----------------|-------------|
| Best Local Similarity | 21.2%; | Pred. No. 4.6e-05;  |                 |             |
| Matches               | 126;   | Conservative 78;  | Mismatches 244; | Indels 146; |
|                       |        |   |                 | Gaps        |
| Qy                    | 24     | GGAYPPAARASAGS-----LRAYVHMTPPSGWLCDPQRPVTHHGAYQLYYLH-----SD 73      |                 |             |
| Db                    | 42     | GLRVDEAAAGCFPWSNEMLOWRSGYHFQAKNYMSDPNCLMYNGWYHMFQYNPVCGTD 101       |                 |             |
| Qy                    | 74     | QNNCGGGNDHASTYDGVAFTHGHTGYMPLRDPFPW-----SGSAVV-----GPA-----NTA 120  |                 |             |
| Db                    | 102    | WDDGQ- EWGHAVSRNLVW-----RTLPIAMVADQWIDILGVLSGSMVLPNGTIVIMYTG 156    |                 |             |
| Qy                    | 121    | GFGAGAV-----VALATOPTDGV-----RKYCEQYLYKSTDGGFTTALPDFVIVNTDGRAA 172   |                 |             |
| Db                    | 157    | ATNASAVEVOCIATPADPNDFLRWTKHPANPVIWSPG-----IGTKD----- 201            |                 |             |
| Qy                    | 173    | TTPAEIENAEWFRDPKTHW-DTARGEWCVCVIGRL-----RYAIFYTSPNLRDWTLRR 223      |                 |             |
| Db                    | 202    | -----FRDPTAWYDESDTDWRTLGLSKDDHGDHGDIAIMMYKTKDFLNYELIP 250           |                 |             |
| Qy                    | 224    | NFYDPNHALGIECPDLFEI--TADQGRHW--VLAASMDAYGIGLPMYAYWT-CTWDG 278       |                 |             |
| Db                    | 251    | GILHRVORTGEWECIDFYPVGHRSDNDSSEMLHVLKASMD-----DERHDYYSLGTYDS 304     |                 |             |
| Qy                    | 279    | EQFHADDLTPOWLDWG-----WDW-----YAAVTPWPSIDAPETKRLAIAMNNWXYAARDVPT 331 |                 |             |
| Db                    | 305    | AANAATPIDP-LDLGILGRYDMGKFYASTSF--YDPAKKRRVLMGYGEVDSKRAV-- 359       |                 |             |
| Qy                    | 332    | DASDGYNGQNSIVRELRLARQPGWYTLTSTPVAALTNYVATATTLTDPDRTDVGSAVLP-- 389   |                 |             |
| Db                    | 360    | --VKGWASIOSVPRITALDEKTR--TNLLLPWVEETLRLNATELSDVTLNCGSVIHPI 415      |                 |             |
| Qy                    | 390    | -WNGRAYEIE-----LDIAWTA TN--VGISVGRSPDGRHTNIGKYGADLYVDRGPSDL 441     |                 |             |
| Db                    | 416    | LROCTQUDIEATHLDAASAVALNEADVGNCSSGGAVNRGALGPFGL-LVLAAG--DR 472       |                 |             |



QY 442 AGYSLAPYSRAAPIDPG-----ARSHLRILVD 470  
Db 473 RGEOTAVFYVSRGLDGLHTSFCDELSSRAKDVTKRVIGSTVPVLDDGEAFSMRLVD 532  
QY 471 TQSVVEFVNAGHTVLSQQVHFAEGDTGISLY-----TDGPAHFTGIVVREIGQA 520  
Db 533 HSIQGFAMGGRTTMTSRYPMEAOEAKVYLFNATGASVMAERLVVHEMDSA 586  
RESULT 12  
US-10-003-392-4  
; Sequence 4, Application US/10003392  
; Patent No. US20020170086A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Caimi, Perry G.  
; APPLICANT: Stoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BB1463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003,392  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 4  
; LENGTH: 609  
; TYPE: PRT  
; ORGANISM: Parthenium argentatum Grey  
US-10-003-392-4  
Query Match 5.4%; Score 155; DB 9; Length 609;  
Best Local Similarity 20.9%; Pred. No. 0.00012;  
Matches 117; Conservative 67; Mismatches 193; Indels 184; Gaps 31;  
QY 40 RAVYHMTPPSGWLCDDPQRPVTHGAYQLYLYHSDQNN--GPGGWDHASTTGDGVAETHHGT 97  
Db 81 RTAFHFQPAKNIYDPNGQMYMGVHFLFYQNPYAPVWGNMWSGHSVSKDMINMY---- 136  
QY 98 VMLPRDFPWSGSVAVGTA--NTAGFAGAVVALATQTDGVRKYOEQYLYWSDGGTT 155  
Db 137 -----ELP-----AIVPTWYDIEGVLSGTVLPNG-----QIFALYTGANDFS 178  
QY 156 ---FTALP-----DPVIVN---TDCRAATTAEIENAEWERDPKTHWDTFARGEWCVICR 204  
Db 179 QLOKRAVPNSDPLLVWVKVEDNPILYTPPGI--GLKYRDPSTVWTGPDGKHRMIMGT 237  
QY 205 LR-----YAAFTYSPNLRDWTLRRNFDYPNHALGGI---ECPDLFEITADD----- 247  
Db 238 KRGNTGMILVYHTT---DYTNVEMLENEPMHSPVNTDMWECVDFYPSVLTNDSALDIAAY 293  
QY 248 --GTRHWVLAASMDAYGICLPMTYAYWTGTWGDGEFHADDLTPQ--WLDWG-----WDWYA 299  
Db 294 GSGIRH-VIKESWEGYGMDF-----YSICTYDA---FNDKWTPDPNPELDVGIGLCRCYGR 344  
QY 300 AVTWFSIDAPETKRLAIAMNNWKYAADVPDTSADGYNQNSIVRELRLARQPGGWYTL 359  
Db 345 PFASKSIEDPVKKRI-----TWAVYGE-----SD--NADDLSR-----GWATV 382  
QY 360 LSTPVAALTNVYATTTLPDRVDCGSVLPW-----NGRAY-EIELD-----IA 402  
Db 383 -----YNVGRTIVLDRKT-GTHLLHWPEVIESLRVNGQFEKEKLEPGSIAPLD 431  
QY 403 WDTATNVGI-----SVGRSPDGTTRHNICKYAGDLYVDRGPSDL 441  
Db 432 IGTAQOLDIVAFKPVDEAALNATSETDNNFACTTSSGAVERSGLGPFGLAVLADGTLSEL 491  
QY 442 AGYSLAPYSRAAPIDPGARSVH-----LRLIV 469  
Db 492 T-----PVYFYIAKADGGV-STHFCTDKLRSLDFDKERVVYGVSTVPVLDDEELTMRLIV 546

QY 470 DTQSVVEFVNAGHTVLSQQVH 490  
Db 547 DHSVVVEAFAGGRIATSRVY 567  
RESULT 13  
US-09-990-385-11  
; Sequence 11, Application US/09990385  
; Publication No. US20020192771A1  
; GENERAL INFORMATION:  
; APPLICANT: Koji YANAI et al.  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PROD  
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE V  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/990,385  
FILING DATE: 10-Sep-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/142,623  
FILING DATE: September 10, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 2001-1611  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 565 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: No. US20020192771A1 relevant  
TOPOLOGY: No. US20020192771A1 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Penicillium roqueforti IAM7254  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..565  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-09-990-385-11  
Query Match 5.2%; Score 147.5; DB 9; Length 565;  
Best Local Similarity 21.8%; Pred. No. 0.00046;  
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;  
QY 26 AVPPAARASAPGSL-----RAVYHMTPPSGWLCDDP-----QRPVTHGAYQLYLYHSDQNN 77  
Db 11 SAPPNLTSLANSLFKTWPRAPRAHLLPPSGNIGDPCGHYTDPKT--GLFHWGLYS-----G 64  
QY 78 PGGWDHASTTGDGVAETHHGTVMPLRPDPFVWGSASVVGTTAN--TAGFGAGAVV--ALATGP 134  
Db 65 ISG-----ATTDDLIV-----TYKDLNPD---GAPSIVAGGKNDPLSVFDGSIPIGIDGMP 112  
QY 135 T-----DGVRYKOEQYLYWSDGGTFTTALPDPVIVNTGAAATTPAEIE 179  
Db 113 TLLYTSVSVLPDHWISPYTRGSETQSLAVSYDGGHFTKL-----NOGPVITPPPAL 165  
QY 180 NAEWFRDPKIHV-----DTARGEWCVICG-----RLRYAAFTYSPNLRD 218

Db 166 NVTAFRDPYVFQSPILDKSVNSTQGTWYVAISGGVHGVGPCQFLYRQNDADFQYWEYLQ 225  
QY 219 W---TLRRNFDPYVFNHALG---GTECPDLPEITADGTRHWVLAASMDAYGIGLP----- 266  
Db 226 WKEPLNTTWCKGDWAGGWFNEVGVNFSNAGVSEDEGEIITICAGESGLPIVPOVS 285  
QY 267 ---MTYAYWTGWDGEGQFHADDLTPQLMDGWDWYAAV---TWPSIDAPETKR-----L 314  
Db 286 SIRDMLVNTGNVTNDGVSFTKPTMAGV-LDVGVSAYAAAGKILPASSQASTKSGAPDREI 344  
QY 315 AIAMNNWKY-AARDVPTDASDYNQNSIVREL-----RLARQ 352  
Db 345 SYVWLTGDLFEQVKGFPPT-AQQWNTGALLPRELNVRTISNVVDNELSRESLTSWRVARE 403  
QY 353 PGGWYLLSPVVAALTYVTATTT-----LPDRVTGSAVLPWN---GRAYEIELDIW 403  
Db 404 DSCQIDLEWGISISRETSYALTSGSFVSGKTLNAGAVPNTSPSSKFFVLTANISF 463  
QY 404 DT-ATNVGISVGRSPDGRHTNIGKY-----GADLYVDRGPSDIAGYSLAPYSRAAAPI-- 456  
Db 464 PTSARDSGIOAGFOVLSSLSLESTIYQFSNESIIVDRSNTSAA-----ARTTAGILS 516  
QY 457 ---DPG-----ARSVHLRLIVDTQSVFVN 479  
Db 517 DNEAGRRLFDVLRNGKEQVETLELTIVVDNSVLEYAN 555

RESULT 14

US-09-990-385-1  
; Sequence 1, Application US/09990385  
; Publication No. US20020192771A1  
; GENERAL INFORMATION:  
; APPLICANT: KOJI YANAI et al.  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCTION  
; OF BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARIANTS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/990,385  
; FILING DATE: 10-Sep-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/142,623  
; FILING DATE: September 10, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee Cheng  
; REGISTRATION NUMBER: 40,949  
; REFERENCE/DOCKET NUMBER: 2001-1611  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acid residues  
; TYPE: amino acid  
; STRANDEDNESS: No. US20020192771A1 relevant  
; TOPOLOGY: No. US20020192771A1 relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Microorganism: Aspergillus niger ACE-2-1  
; (ATCC 20611)

FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..635  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-990-385-1  
  
Query Match 5.0%; Score 141.5; DB 9; Length 635;  
Best Local Similarity 19.1%; Pred. No. 0.0017;  
Matches 125; Conservative 83; Mismatches 245; Indels 203; Gaps 28;  
  
QY 26 AVPAARASAPGSL-----RAVYHMTPPSWLWCDPQRPVT--THGAYQLYLLISDQNGNP 78  
Db 8 APPPTLNLSTPLNTLHFVWPRRAHILPAEQIGDPCAHYTDPTSLGFLHVGFLH--DGDGI 65  
QY 79 GGWDHAS-----TTDCGVAFTTHGCTVMPLRPDPF---VMSGSAV-VGTANTAGFCAGAVVA 129  
Db 66 AGATTANLATYTTDSNGSP---LIQPGKNDPVAVDGAVIPGVNNTPTLLYTSVSF 121  
QY 130 LATQ-PTDGVYKQEQYLYWSTDGFTF-----TALPDP----- 162  
Db 122 LPIHWSIPYTRGSETQSLAVARDGGRFDKLDQGPVIADHPFAVDVTAFRDPPEVFRSAKL 181  
QY 163 -VIVNTDGRAATTPAEIENA-EWFRDPKIHWDTARGEWCVIGRLRYAAFTYSPNLRD-- 218  
Db 182 DVLLSLDEEVARNETAQQAVDGVWTEKNAPYVAVSGVGVHGVGPAOFLYRONGGNASEFQ 241  
QY 219 -----WTLRRNFDPNHALGGI-----ECPDLFEITADDT 249  
Db 242 WYELGEMWQEQATNSSWDEGTWAGWGF---NFETGNVLFLEEGHDPQOTGEVFTLGT 298  
QY 250 RHWVLAASMDAYGIGLPMT-----YAYWTGTWGDGEQFHADDLTPQ---WLDWGW 295  
Db 299 E-----CSGLPIVPOVSSIHDMLWAAGEVGVGSEGEAKVFEFSPSMAGFLDWGE 347  
QY 296 DWTAAV--TWPSIDAE-----TKRLATAMNNWKYAAARDVPTDASDYGNGNSIVRE 346  
Db 348 SAYAAAAGKVLPAASAVSKTSGVGVDRYVFWLTGQOYEQADGFPPTAQQGWTSLLPRE 407  
QY 347 LR-----LAKQPG-----GWYTLSTPVAALTYVTATTT 377  
Db 408 LKQVTENVVDNELVREEGSVWVGESDQNTARLTIGITIARETKAALLAN--GSVTAE 465  
QY 378 PDRTVGSAVLPW---NGRAYEIELDIADWTATN-----VGISVGRSPDGRHT 423  
Db 466 EDRTLQATAVVPFAQSPSSKFFVLTALQLEPPASARSPLQSGFEILASELERTAIYYQFS 525  
QY 424 NICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPG-----AR 461  
Db 526 N-----ESLVVDRSQTs-----AAAPNGLDSTESGKRLFDVIENGQEQVE 569  
QY 462 SVHLRLIVDTQSVFVNAGHTVLSQOVHFAEGDTGISLYTDG-GPAHFTGIYVRE 516  
Db 570 TLDLTVVVDNAVVEVYANGRFALSTWARSWYDNTQIRFFHNGEVEGVQFRNVSVSE 625

RESULT 15

US-09-990-385-13  
; Sequence 13, Application US/09990385  
; Publication No. US20020192771A1  
; GENERAL INFORMATION:  
; APPLICANT: KOJI YANAI et al.  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCTION  
; OF BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARIANTS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/990,385  
FILING DATE: 10-Sep-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/142,623  
FILING DATE: September 10, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 2001-1611  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 574 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: No. US20020192771A1 relevant  
TOPOLOGY: No. US20020192771A1 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Scopulariopsis brevicaulis IFO4843  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1...574  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-990-385-13

Query Match 4.9%; Score 140.5; DB 9; Length 574;  
Best Local Similarity 21.9%; Pred. No. 0.0018;  
Matches 129; Conservative 56; Mismatches 245; Indels 159; Gaps 29;

QY 26 AVPPAARASAPGSL----RAVYHMTPPSGWLCDPQRPVT--THGAYQLYLHSDQNNPG 79  
Db 19 SAPPNLSTLANNLETPWRAHVLPPQNOIGDCPMHYTDPETGIFHVGWLYN--GNGAS 76  
QY 80 GMDHASTDGVAFTHGTVMPLRPDPFVWMSGAVV---GTANTAGFGAGAVV--ALATOP 134  
Db 77 G---ATTEDLV-----TYQDLNPD-----GAQMILPGGVNDPLAVFDGAVIPSGIDGKP 122  
QY 135 TDGVRKYQEQLYV-----STDGGFTTAL-PDPYIVNTDGRAATP---- 175  
Db 123 TMNYSVSYMPISWISYATRGSETHSLAVSDGGKNTKLVQGPVTPSPFPGANVTSWRD 182  
QY 176 -----AEIENAEWF-----RDPKIHWDTARGEWVCVIGR 204  
Db 183 PFLFQNPQFDSLESENGTWYTVISGGIHGDCPSAFLYRQHDPDFQYWEYLGFWNNEGN 242  
QY 205 LRYAAYTSPNLRDWTLRNFDYPNHALGGIE----CPDLFEITADDGTRHWVLAASMDA 260  
Db 243 STWGS-----GDWAGRWGYNFEVINIVGLDDGDGYNPD-GEIFATVGT-EWSFDPKPKQ 293  
QY 261 YGIGLPMYAYWTGTWDCGEQFHADDLTPOWLDGWDAVAV--TWPSIDAPETKR----- 313  
Db 294 ASDNREMLWAGNMTLEDGDIKFTPSMAGYLDWGLSAYAAAGKELPASSKPSOKSGAPDR 353  
QY 314 -LAIAMNNWYKARDVPTDASDYGNGQNSIVREL-----RLAROPGGWYTLIS 361  
Db 354 FVSYLWLTGDFEGHDFPT-POONTGSLLLPRELSVGTIPNVVDNELARETGSW--RVG 410  
QY 362 TPVAALTNVT-----ATTLPDRTVGDS-----AVLPWNGRAYEIELDIAMDATNVG 410  
Db 411 TNDTGVLELVLTQEIARETLAEMTSGNSFTEASRNSSPGSTAFQOSLDSKFFVLT-AS 469  
QY 411 ISVGRSPDGRTRTNICKYGADLYVDRGPDLAGYSLA-----PYSRAAPIDPG---- 459  
Db 470 LSF---PSSARDSL-KAGFEILSEFESTTVYQFNSIIIDRSNSAAALTTDGDIT 525

Search completed: May 6, 2003, 12:44:10  
Job time : 21.5337 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:32:47 ; Search time 58.6125 Seconds  
(without alignments)  
1515.790 Million cell updates/sec

Title: US-09-868-328B-1  
Perfect score: 2845  
Sequence: 1 MTPAISRRVLOGAGAGALA.....TDGPAHFTGIVVREIQAI.521

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 785308.seqs, 170526417 residues

Total number of hits satisfying chosen parameters: 785308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 2845  | 100.0       | 521    | 5  | US-09-868-328A-2    |
| 2          | 2845  | 100.0       | 521    | 5  | US-09-868-328B-1    |
| 3          | 2837  | 99.7        | 519    | 5  | US-09-868-328B-4    |
| 4          | 944   | 33.2        | 516    | 6  | US-10-369-493-23310 |
| 5          | 556   | 19.5        | 677    | 6  | US-10-369-493-23212 |
| 6          | 449.5 | 15.8        | 541    | 6  | US-10-369-493-13045 |
| 7          | 359.5 | 12.6        | 532    | 6  | US-10-369-493-22074 |
| 8          | 348   | 12.2        | 581    | 6  | US-10-369-493-22749 |
| 9          | 313.5 | 11.0        | 311    | 6  | US-10-369-493-10616 |
| 10         | 304   | 10.7        | 670    | 1  | PCT-US03-07858-1502 |
| 11         | 304   | 10.7        | 670    | 6  | US-10-289-757-176   |
| 12         | 304   | 10.7        | 670    | 6  | US-10-389-566-1502  |
| 13         | 295   | 10.4        | 490    | 6  | US-10-369-493-3411  |
| 14         | 294   | 10.3        | 668    | 6  | US-10-289-757-87    |
| 15         | 291   | 10.2        | 677    | 1  | PCT-US03-07858-1977 |
| 16         | 291   | 10.2        | 677    | 6  | US-10-389-566-1977  |
| 17         | 285.5 | 10.0        | 583    | 6  | US-10-289-757-83    |
| 18         | 283.5 | 10.0        | 583    | 6  | US-10-289-757-79    |
| 19         | 282   | 9.9         | 662    | 1  | PCT-US03-07858-789  |
| 20         | 282   | 9.9         | 662    | 6  | US-10-389-566-789   |
| 21         | 279   | 9.8         | 636    | 1  | PCT-US03-07858-1296 |
| 22         | 279   | 9.8         | 636    | 6  | US-10-389-566-1296  |
| 23         | 279   | 9.8         | 646    | 1  | PCT-US03-07858-1505 |
| 24         | 279   | 9.8         | 646    | 6  | US-10-389-566-1505  |
| 25         | 268.5 | 9.4         | 673    | 1  | PCT-US03-07858-482  |
| 26         | 268.5 | 9.4         | 673    | 6  | US-10-389-566-482   |

|    |       |     |     |   |                     |                   |
|----|-------|-----|-----|---|---------------------|-------------------|
| 27 | 267.5 | 9.4 | 577 | 1 | PCT-US03-07858-2157 | Sequence 2157, Ap |
| 28 | 267.5 | 9.4 | 577 | 6 | US-10-389-566-2157  | Sequence 2157, Ap |
| 29 | 267.5 | 9.4 | 655 | 1 | PCT-US03-07858-2432 | Sequence 2432, Ap |
| 30 | 267.5 | 9.4 | 655 | 6 | US-10-389-566-2432  | Sequence 2432, Ap |
| 31 | 267.5 | 9.4 | 663 | 1 | PCT-US03-07858-1624 | Sequence 1624, Ap |
| 32 | 267.5 | 9.4 | 663 | 6 | US-10-389-566-1624  | Sequence 1624, Ap |
| 33 | 266.5 | 9.4 | 477 | 6 | US-10-369-493-8940  | Sequence 8940, Ap |
| 34 | 266   | 9.3 | 648 | 1 | PCT-US03-07858-1489 | Sequence 1489, Ap |
| 35 | 266   | 9.3 | 648 | 6 | US-10-389-566-1489  | Sequence 1489, Ap |
| 36 | 261   | 9.2 | 647 | 1 | PCT-US03-07858-1594 | Sequence 1594, Ap |
| 37 | 261   | 9.2 | 647 | 6 | US-10-389-566-1594  | Sequence 1594, Ap |
| 38 | 260.5 | 9.2 | 640 | 1 | PCT-US03-07858-1466 | Sequence 1466, Ap |
| 39 | 260.5 | 9.2 | 640 | 6 | US-10-389-566-1466  | Sequence 1466, Ap |
| 40 | 260.5 | 9.2 | 669 | 1 | PCT-US03-07858-788  | Sequence 788, App |
| 41 | 260.5 | 9.2 | 669 | 6 | US-10-389-566-788   | Sequence 788, App |
| 42 | 258.5 | 9.1 | 490 | 6 | US-10-417-884-5048  | Sequence 5048, Ap |
| 43 | 256.5 | 9.0 | 673 | 1 | PCT-US03-10544-21   | Sequence 21, Appl |
| 44 | 256.5 | 9.0 | 673 | 6 | US-10-409-701-21    | Sequence 21, Appl |
| 45 | 255   | 9.0 | 597 | 6 | US-10-289-757-81    | Sequence 81, Appl |

ALIGNMENTS

RESULT 1

US-09-868-328A-2  
; Sequence 2, Application US/09868328A  
; GENERAL INFORMATION:

; APPLICANT: RHEE, Sangki

; APPLICANT: SONG, Kibang

; APPLICANT: KIM, Chulho

; APPLICANT: RYU, Eunja

; APPLICANT: LEE, Yongbok

; TITLE OF INVENTION: Enzymatic Production of Diffructose Dianhydride IV from sucrose

; TITLE OF INVENTION: relevant enzymes and genes coding for them

; FILE REFERENCE: 24679

; CURRENT APPLICATION NUMBER: US/09/868,328A

; CURRENT FILING DATE: 2001-06-18

; PRIOR APPLICATION NUMBER: PCT/KR00/01183

; PRIOR FILING DATE: 2000-10-19

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 521

; TYPE: PRT

; ORGANISM: Arthrobacter ureafaciens

US-09-868-328A-2

Query Match

Best Local Similarity 100.0%; Score 2845; DB 5; Length 521;

Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |     |   |     |
|----|-----|---|-----|
| QY | 1   | MTPAISRRVLOGAGAGALIFGGAVPPAARASAPGSLRVAVHMTPPSGWLCDDPDPVT   | 60  |
| DB | 1   | MTPAISRRVLOGAGAGALIFGGAVPPAARASAPGSLRVAVHMTPPSGWLCDDPDPVT   | 60  |
| QY | 61  | THGAYQLYLHSDNNGPGWDHASTTDCGVAFTHGCTVMPRLPDPFPVWSGSAVVGTA    | 120 |
| DB | 61  | THGAYQLYLHSDNNGPGWDHASTTDCGVAFTHGCTVMPRLPDPFPVWSGSAVVGTA    | 120 |
| QY | 121 | GFAGAVVALATQPTDGVKRYOEQYLYWSTDDGGFTTALPDVIVNTDGRAATPAEIN    | 180 |
| DB | 121 | GFAGAVVALATQPTDGVKRYOEQYLYWSTDDGGFTTALPDVIVNTDGRAATPAEIN    | 180 |
| QY | 181 | AEWFRDPKTHWDTARGEWCVIGRLRYAAYTSPNLRDWTLRNFDPYPNHALGGIECPDL  | 240 |
| DB | 181 | AEWFRDPKTHWDTARGEWCVIGRLRYAAYTSPNLRDWTLRNFDPYPNHALGGIECPDL  | 240 |
| QY | 241 | FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTWDGQFHADDLTPOWLDGWDWYAA   | 300 |
| DB | 241 | FEITADDGTRHWVLAASMDAYGIGLPMYAYWTGTWDGQFHADDLTPOWLDGWDWYAA   | 300 |
| QY | 301 | VTWPSIDAPETKRLATIAWNNMKYAAARDVPTDASDGYNGQNSIVRELRLARPGGWYTL | 360 |

Db 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTL 360  
QY 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420  
Db 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420  
QY 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480  
Db 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480  
QY 481 GHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREIGQAI 521  
Db 481 GHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREIGQAI 521

## RESULT 2

US-09-868-328b-1  
; Sequence 1, Application US/09868328B  
; GENERAL INFORMATION:  
; APPLICANT: RHEE, Sang-Ki  
; APPLICANT: SONG, Ki-Bang  
; APPLICANT: KIM, Chul-Ho  
; APPLICANT: RYU, Eun-Ja  
; APPLICANT: LEE, Yong-Bok  
; TITLE OF INVENTION: Enzymatic production of difructose dianhydride IV from sucrose and  
; FILE REFERENCE: 24679  
; CURRENT APPLICATION NUMBER: US/09/868,328B  
; CURRENT FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: PCT/KR00/01183  
; PRIOR FILING DATE: 2000-10-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 1  
; LENGTH: 521  
; TYPE: PRT  
; ORGANISM: Arthrobacter ureafaciens K2032  
US-09-868-328b-1

Query Match 100.0%; Score 2845; DB 5; Length 521;  
Best Local Similarity 100.0%; Pred. No. 2.7e-209; Mismatches 0; Indels 0; Gaps 0;  
Matches 521; Conservative 0;  
QY 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPORPVT 60  
Db 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPORPVT 60  
QY 61 THGAYOLYLHSDQNGPGGDHASTTDGVAFTTHGCTVMPLRDPDPVWGSAGVVGTA 120  
Db 61 THGAYOLYLHSDQNGPGGDHASTTDGVAFTTHGCTVMPLRDPDPVWGSAGVVGTA 120  
QY 121 GFAGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDVIVNTDGRAATTAEIEN 180  
Db 121 GFAGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDVIVNTDGRAATTAEIEN 180  
QY 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
Db 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
QY 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDVAA 300  
Db 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDVAA 300  
QY 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTL 360  
Db 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTL 360  
QY 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420  
Db 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420  
QY 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480

Db 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480  
QY 481 GHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREIGQAI 521  
Db 481 GHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREIGQAI 521  
RESULT 3  
US-09-868-328b-4  
; Sequence 4, Application US/09868328B  
; GENERAL INFORMATION:  
; APPLICANT: RHEE, Sang-Ki  
; APPLICANT: SONG, Ki-Bang  
; APPLICANT: KIM, Chul-Ho  
; APPLICANT: RYU, Eun-Ja  
; APPLICANT: LEE, Yong-Bok  
; TITLE OF INVENTION: Enzymatic production of difructose dianhydride IV from sucrose  
; FILE REFERENCE: 24679  
; CURRENT APPLICATION NUMBER: US/09/868,328B  
; CURRENT FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: PCT/KR00/01183  
; PRIOR FILING DATE: 2000-10-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 4  
; LENGTH: 519  
; TYPE: PRT  
; ORGANISM: Arthrobacter ureafaciens K2032  
US-09-868-328b-4

Query Match 99.7%; Score 2837; DB 5; Length 519;  
Best Local Similarity 100.0%; Pred. No. 1.1e-208; Mismatches 0; Indels 0; Gaps 0;  
Matches 519; Conservative 0;  
QY 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPORPVT 60  
Db 1 MTPAISRRVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDPORPVT 60  
QY 61 THGAYOLYLHSDQNGPGGDHASTTDGVAFTTHGCTVMPLRDPDPVWGSAGVVGTA 120  
Db 61 THGAYOLYLHSDQNGPGGDHASTTDGVAFTTHGCTVMPLRDPDPVWGSAGVVGTA 120  
QY 121 GFAGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDVIVNTDGRAATTAEIEN 180  
Db 121 GFAGAGAVVALATOPTDGVVRKYQEOYLWSTDGGFTTALPDVIVNTDGRAATTAEIEN 180  
QY 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
Db 181 AEWFRDPKTHWDTARGEWCYVIGRLRYAAYTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
QY 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDVAA 300  
Db 241 FEITADDGTRHWVLAASMDAYGIGLPMTYAYWTGTWDSQFHADDLTPQWLWDGWDVAA 300  
QY 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTL 360  
Db 301 VTWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTL 360  
QY 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420  
Db 361 STPVAALTNYVTATTTLPDRTVDSAVLPWNGRAYEIELDIANDTATNVGISVGRSPDGT 420  
QY 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480  
Db 421 RHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVEFVNA 480  
QY 481 GHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREIGQ 519  
Db 481 GHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREIGQ 519

## RESULT 4

US-10-369-493-23310  
; Sequence 23310, Application US/10369493  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 23310

; LENGTH: 516  
; TYPE: PRT

; ORGANISM: Bacillus subtilis  
US-10-369-493-23310

Query Match 33.2%; Score 944; DB 6; Length 516;  
Best Local Similarity 41.8%; Pred. No. 8.6e-64;  
Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAYVHTPPSGWLCDDPQRPVTHGAYQLYLSHSDQNN--GPGGDHASTTDGVAFTTHGCT 97

DB 44 RAYHTTDPKKNNDQKQIYFDGKYHYFLYNRDYPKNGTEWRHVAVEDLVHWTDEGV 103

QY 98 VMP--LRPDPFVWSSGAVVGTANTAGAGAVVALATOPTDGVYKQEOYLWSTDGGFT 155

DB 104 AIPKVTNPDGDIWTSVVVDKENTAGFGKNALVAIVTQPSAKDKK-QEOYLWSTDKGS 162

QY 156 F-----TALPDVIVNTDGRAATTPAEIENAEFRDPKIHWDTARGEWCVCVIGRLRYAA 209

DB 163 FKFYSGNPVMPNP---GTDD-----FRDPKVIWDDQDNKVMVMAEGSKIG 205

QY 210 FYTSNLRDWTLLRRNFDYPNHALGGIECPDLFEITADGTRHWLAAASMDAYGIGLPMY 269

DB 206 FYESNLKDNHYSYSGF-FPEQA-GWVECDPLMMRASDGTNRKWLGSANGKPMGKPNY 263

QY 270 AYWTGTWGEQFADDLTPQLWDGWDWYAAVTPSIDA--PETKRLATAWNNNKKYAAAR 327

DB 264 AWTGSFDCGKFKADQTEAQWLDYGFWDYGGVTFEEDSKSDTDPLEKRYALAWNNNDY-AN 322

QY 328 DVPTDASDGYNGQNSIVRELRLAROPGQWYTLSTPVAALTYVATTTLPRTVDGSAV 387

DB 323 NTPT-MKNGFNGTDSVIRELRKEQ-DGTYSLSVQPIEALEQLTVSTDEIQDQVNGSKT 380

QY 388 LPWNGRAYEILDIAWDATNNGISVGRSPDGTNRHNTICKY--GADLYVDKRGPSDLAGYS 445

DB 381 LSIITGDTYOLDTDLSELKNAAGVRLRESEDKRHDIDVGFAGGAYVYVNRATNQPDKS 440

QY 446 LAPYSRAAAPIDPGARSVHLRLIVDTQSVEFVFNAGHTVLSQQVHFAEGDGTGSLYTDGG 505

DB 441 -NTYVESKAPYDVNKKVHLKILVDKTTIEVFVGDKTVFSNEVPKPKEDKGITLYSDGG 499

QY 506 PAHFTGIVVR 515

DB 500 TASFRKNITVK 509

## RESULT 5

US-10-369-493-23212

; Sequence 23212, Application US/10369493  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 23212

; LENGTH: 677  
; TYPE: PRT

; ORGANISM: Bacillus subtilis  
US-10-369-493-23212

Query Match 19.5%; Score 556; DB 6; Length 677;  
Best Local Similarity 32.2%; Pred. No. 6.3e-34;  
Matches 160; Conservative 74; Mismatches 217; Indels 46; Gaps 18;

QY 40 RAYVHTPPSGWLCDDPQRPVTHGAYQLYLSHSDQNN--GPGGDHASTTDGVAFTTHGCT 97

DB 35 RPQYHFTPEANWMDNPGMVYAGYHLEFYQYHPYGLQWPMHGHAVSKDLVTWEH--L 92

QY 98 VMPLRPD--FPVWSSGAVVGTANTAGFGAG---AVVALATOPTDGVYKQEOYLWSTDG 152

DB 93 PVALYDPDEKGTIFSGSAVVDKXNNTSGFTGKEKPLVAIYTDREG---HGVQSIAYSNDK 149

QY 153 GFTFTALP-DPVTIVNTDGRAATTPAEIENAEFRDPKIHWDTARGEWCVCVIGRLRYAAFY 211

DB 150 GRITWKYAGNPVLPN-----PGKDK---FRDPKVFVWEKEKKWVAVLAAGDRILLY 197

QY 212 TSNPLRDLTLLRRNFDYPNHALGGI-ECPLFEITAD--DGTRHVLAAASMDAYGIGLPMY 268

DB 198 TSKNLKQWYASEFGDQSGHGVWECPLFELPVDGPNQKWKVMQVSGVNGAVSGSG 257

QY 269 YAYWTGTWGEQFADDLTPQ--WLDWGDWYAAVTPSIDAPETKRLATAWNNNKKYAA 326

DB 258 MOYFVGDFDGTGTHFKNENPNKVLWTDYGRDYAAVSWSDIPSTDSRRLWLGWMSNQY-A 316

QY 327 RQVPTDASDGYNGQNSIVRELRLAROPGQWYTLSTPVAALTYVATTTLPRTVDGSA 386

DB 317 NDVPTSP---WRSAYSIIPRELKKAFTGEG-VRVVOPTPVKELETIRGTSKKKNLITSPAS 372

QY 387 ---VLPWNGRAYEYI--ELDIAMDITATNNGISVGRSPDGTNRHNTICKYGADLYVDGRPSDL 441

DB 373 HNVLAGQSGDAYEINAEFKVSPGSAAEFGFKVKTGENQFTKVDYDRNNAKLFVDRSECN 432

QY 442 AGYSLA-PYSRAAAPIDPGARSVHLRLIVDTQSVEFVFNAGHTVLSQQVHFAEGDGTGSL 500

DB 433 DTENPAFNTGKETAPLKPVGKVKLRFYVDRSSVEFVNGDKGQVITDIILPDRSSKGLEL 492

QY 501 YTDGGFAHFTGIWVREI 517

DB 493 YAANG-----GVKVKSL 504

## RESULT 6

US-10-369-493-13045

; Sequence 13045, Application US/10369493  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 13045

; LENGTH: 541  
; TYPE: PRP  
; ORGANISM: Aspergillus nidulans  
US-10-369-493-13045

Query Match 15.88; Score 449.5; DB 6; Length 541;  
Best Local Similarity 28.1%; Pred. No. 6.5e-26;  
Matches 160; Conservative 77; Mismatches 205; Indels 127; Gaps 29;  
QY 40 RAVYHMTTPSGWLCDDPQPVTHGAYLYLHSDONNGPGG-----WDHASTTD---- 88  
Db 6 RPQYHFTPAQWINDPNGLYHNGTYHMFYQYN-----PGVAHEAMSWGHATSDQLTHW 60  
QY 89 ---GVAETHGTVMPLRDPFP-----VWSSAVVGTANTAGFG-----AGAVVALAT- 132  
Db 61 DERPVALTARG-----FPGKNISQEFSSGSAVDVDNSSGYGNGMGTAAPWAMVYTS 112  
QY 133 -----QPTDGVK-----YQEQLYLWSTGGFTTFA--L 159  
Db 113 FVCYPSLLIEWSDMLVLTCKVCLGQTLPSGKKVRASQOSQSLAYSLDSGETWTDTDDA 172  
QY 160 PDVIVNTDGRAATTAEIENAEWFRDPKTHWDTARGEW--CVIGRLRYAAFTSPNLR 217  
Db 173 VNPILQ-----PPHPYESQYKEW-RDPNVFVHAPKWKILTTALSDHLKLLIYTSTDLK 226  
QY 218 DWTLRNRPDHALGGI-ECPLDFTADDGTRH--WVLAASMDAYGIGLPMTYA---- 270  
Db 227 NWTLASEFG-PYNAVGVWECSPFPLPVDNDESNNKVAIIGLNP--GPPCTGVSGNQ 283  
QY 271 YWTGTWDEGF--HADDLTQ-----WLDGWWDWYAAVTFWSDAPETKRLAIAMNNWKY 324  
Db 284 YILQFNGHFLPDAESLHGEGRANWLDYDGFYALVYNGL--PEQRTVIAMSMNQY 341  
QY 325 AARDVPTDASDGYNGQNSIVRELRL-----AROP-GGWYTLTSTPVAALTNVYAT 374  
Db 342 AEK-VPTTV---WRNANTIPRLGLKTINGKVLVQVEPEANWAAIMSKP-----QTTRI 391  
QY 375 TTLPDRVDSGAVLPWNGR--AYEILDIAWDTATNVGIVSGRSPDGTNRHNGKYGA-- 430  
Db 392 LSLPANSTRNLGSL--GKTLAANLTFASAGEQSGSFGISLLATQNTTQTLIGYNFATK 448  
QY 431 DLVDRCPGSLAGYSLAPYSRAAPIDPG--ARSVHLRLVDTOSVEVFNAGHTVLSQQV 489  
Db 449 HIFVDRASGNTSDETFASVYRAPLPADDDGTTLRIFADMSVEVFGCGGETTLTAQV 508  
QY 490 HFAEGDTGISLYTDGGPFAHTGIVVREIG 518  
Db 509 FPEGATHARLFANGAGVEDVLLSVSKVG 537

RESULT 7  
US-10-369-493-22074  
; Sequence 22074, Application US/10369493  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO: 22074  
; LENGTH: 532  
; TYPE: PRP  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22074

Query Match 12.6%; Score 359.5; DB 6; Length 532;  
Best Local Similarity 26.0%; Pred. No. 4.9e-19;  
Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;  
QY 19 LALIFGAVPPAARASAPGSLRAVYHMTTPSGWLCDDP-----ORPVTTHGAYLYLHLS 72  
Db 7 LFLLAGFAAKISAMTNETSDRPLVHFTPNKGNWMDPNGLWYDEKDAKH-----LYEQYN 62  
QY 73 DON---NGPCGWPHASTTDCGVAETHG--TVMLPLRDPDPVWMSGAVVGTANTAGFGAGAV- 127  
Db 63 PNDTVNGTPLFWGHATSDDLTNMEDOPIATAPKRNDSGAFSGMVVDYNTSGFFNTID 122  
QY 128 -----VALATQPTDGVKRYQEQLYLWSTGGFTTAL-PDPVIYNTDGRAATTPAETENA 181  
Db 123 PRORCAIWTYNTP---ESEQYISYSLDGGYTFTEYQKNPVL-----ANS 166  
QY 182 EWRDRPKIHWDTARGEWVCVIGRLR--YAAFTYSPNLRDWTLRNFDYPNHALGG--IEC 237  
Db 167 TQRPDRPKVFWYEPQKWMTAASKQDYKIEYSSDDLUKSWKLESF--ANEGLGYOYEC 224  
QY 238 PDLFEITA--DDGTRHNVLAASMDAYGIGLPMTYA---YWTGTWDEGQFHADDLTPOWLD 292  
Db 225 PGLIEVTEQDPSKSYWMEFISINP---GAPAGSFGNQYFVGSFNGTHFEAFDQSRVVD 281  
QY 293 WGDWYAAVTPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRL--- 349  
Db 282 FGKDYALQTFNTDPTYGSALGIAMASNNWEYS-AFVPTNP---WRSSMLSRKFSLNT 337  
QY 350 --ARQPGGWYTLTSTPVAALTN-----YVTATTLTLPDRD--VDGSVLPWNGRAYEIEL 399  
Db 338 YQNPETELINLKAEPILNISAGPWSRFATNTTLTKANSYNVD-----LSN:GTLEFEL 393  
QY 400 DIAWDTATNVGIVSG-----RSPDGTNRHNTIGKYGADLYVDRGSDLAGYSLAPY 449  
Db 394 YVAVNTQTITTSKVFADLSLWFKGLEPDEYLRMGPEVSASSFFELDRGSKVKFVKNPY 453  
QY 450 --SRAAAPIDPGARSVHLRI-----LVDTOSVEVFNAGHTVLSQQVHFAEGD----- 495  
Db 454 FTRMSVNNOPFKSENDSLKYKYVGLDQNLILEYFNDSGDVSTNTYFMTTGNALGSVNM 513  
QY 496 -TGIS--LYTD 503  
Db 514 TTGVDNLFYID 524

RESULT 8  
US-10-369-493-22749  
; Sequence 22749, Application US/10369493  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO: 22749  
; LENGTH: 581  
; TYPE: PRP  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-22749  
Query Match 12.2%; Score 348; DB 6; Length 581;  
Best Local Similarity 24.9%; Pred. No. 4.2e-18;  
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;  
QY 28 PPAARASAPGSL-----RAYHMTTPSGWLCDDPQPVTHGAYLYY 69





Db 415 WGVGETDSRADY-----AKGWASLQSPRTVVLDTKTGS--NLIQWPVVVEVETLRTNST 468  
QY 376 TLPDRVTGDSAVLPWN--GRAYEIE-----LDIAWDATATNVGIVSGRSPDGRHTNI 425  
Db 469 NLGSIIVHSGSVPLSLHRTQDIEASFRLLDPLDVAAKEADVGYNCSTSGGAAGRGAL 528  
QY 426 GKYG-----AD-----LYVDRG-PSDLAGYSLAPYSRAA-----453  
Db 529 GPFGLLVLADARRHGGDEQTAFFVYVARGLDGNLRFHCHDESRSRANDIVKRVVGN 588  
QY 454 APIDPGARSVHLRLTVDTSQVEFVNAGHTVLSQOVHFAE---GDTGISLYTD--GGPAH 508  
Db 589 VPVLGD-EALSVRLVDHSIVESFAQGRSVVTSRVYPTTEAIYANAGVYLFNNATGARVT 647  
QY 509 FTGIVVREI 517  
Db 648 ATSLVHEM 656

## RESULT 11

US-10-289-757-176

; Sequence 176, Application US/10289757

; GENERAL INFORMATION:

; APPLICANT: Demmer, Jeroen

; APPLICANT: Forster, Richard L

; APPLICANT: Gibson, John Bryan

; APPLICANT: Shenk, Michael Andrew

; APPLICANT: Norriss, Geoffrey

; APPLICANT: Glenn, Matthew

; APPLICANT: Saulsbury, Keith Martin

; APPLICANT: Hall, Claire

; TITLE OF INVENTION: Compositions isolated from forage

; TITLE OF INVENTION: grasses and methods for their use

; FILE REFERENCE: 11000.10610

; CURRENT APPLICATION NUMBER: US/10/289,757

; PRIOR FILING DATE: 2002-11-07

; PRIOR APPLICATION NUMBER: 60/337,703

; PRIOR FILING DATE: 2001-11-07

; NUMBER OF SEQ ID NOS: 218

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 176

; LENGTH: 670

; TYPE: PRT

; ORGANISM: Lolium perenne

US-10-289-757-176

Query Match 10.7%; Score 304; DB 6; Length 670;

Best Local Similarity 24.3%; Pred. No. 1.2e-14;

Matches 148; Conservative 74; Mismatches 243; Indels 144; Gaps 26;

QY 4 AISRRVLOQAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVTHG 63  
Db 97 AASESGMLGADAGSNAPFWSNMLQWRTG-----FHFQPEKNMNDPNGPVYKG 148  
QY 64 AYQLYLHSDONNGPG---GWDHASTTDGVAFTTHGTVMLRPDFF-----VMSGSA 112  
Db 149 WYHLFYQYNEGAIGNKTAHGAHSRDLMLRHLPIAM-----FPDQWYDINGAWSGSA 203  
QY 113 VGTANTAGGAGAVALATQPTDGVKRYQEOYLWSTDGGFTTALPDPIVYN---TDG 169  
Db 204 TV-----LPDGRIVMLYTGSTNA--SVQVQCLAFPSDPS-----DPLLTNWTKEYG 247  
QY 170 RAATTPAEIENAEWFRDPKIH--DTARGEWVCVIG-----RLRYAAFTYSPNLRDWTLLR 223  
Db 248 NPVLYPPPHVGEKDFRPTTAWYDGGDMWRIVIGSKDNRNRRAGMALTYKTKNFHFDELVP 307  
QY 224 NFDYPNHALGGICBDFLFEI-----TADDGTRHWLVA 255  
Db 308 GYLHRVPATGMWECIDILYPVGGARGIDMTAEVAAASNSGGGEVLHVWKESSDDDRH---363  
QY 256 ASMDAYGIGLPMYAYWTGTDGCEQFHADDLTPQWLDGWDWYAAVTPSIDAPETKRLA 315  
Db 364 ---DYALG---RYDAATNKNWTPDLADADVIGLRYDNG--KFYASKTF---YDPAKKRRVL 414

QY 316 IAWNNMKNYAARDVPTDASDGYNGQNSIVRELRLAROPGGMWYLLSTPPVAALTNVYVATT 375  
Db 415 WGVGETDSRADY-----AKGWASLQSPRTVVLDTKTGS--NLIQWPVVVEVETLRTNST 468  
QY 376 TLPDRVTGDSAVLPWN--GRAYEIE-----LDIAWDATATNVGIVSGRSPDGRHTNI 425  
Db 469 NLGSIIVHSGSVPLSLHRTQDIEASFRLLDPLDVAAKEADVGYNCSTSGGAAGRGAL 528  
QY 426 GKYG-----AD-----LYVDRG-PSDLAGYSLAPYSRAA-----453  
Db 529 GPFGLLVLADARRHGGDEQTAFFVYVARGLDGNLRFHCHDESRSRANDIVKRVVGN 588  
QY 454 APIDPGARSVHLRLTVDTSQVEFVNAGHTVLSQOVHFAE---GDTGISLYTD--GGPAH 508  
Db 589 VPVLGD-EALSVRLVDHSIVESFAQGRSVVTSRVYPTTEAIYANAGVYLFNNATGARVT 647  
QY 509 FTGIVVREI 517  
Db 648 ATSLVHEM 656

## RESULT 12

US-10-389-566-1502

; Sequence 1502, Application US/10389566

; GENERAL INFORMATION:

; APPLICANT: Monsanto Technology, LLC

; APPLICANT: Laurie, Cathy C

; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants

; FILE REFERENCE: 38-77(52900)D

; CURRENT APPLICATION NUMBER: US/10/389,566

; CURRENT FILING DATE: 2003-03-31

; PRIOR APPLICATION NUMBER: US 60/365,301

; PRIOR FILING DATE: 2002-03-15

; PRIOR APPLICATION NUMBER: US 60/391,786

; PRIOR FILING DATE: 2002-06-25

; PRIOR APPLICATION NUMBER: US 60/392,018

; PRIOR FILING DATE: 2002-06-26

; NUMBER OF SEQ ID NOS: 2459

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1502

; LENGTH: 670

; TYPE: PRT

; ORGANISM: Lolium perenne

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (69)..(69)

; OTHER INFORMATION: Xaa can be any naturally occurring amino acid

US-10-389-566-1502

Query Match 10.7%; Score 304; DB 6; Length 670;  
Best Local Similarity 24.3%; Pred. No. 1.2e-14;  
Matches 148; Conservative 74; Mismatches 243; Indels 144; Gaps 26;

QY 4 AISRRVLOQAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVTHG 63  
Db 97 AASESGMLGADAGSNAPFWSNMLQWRTG-----FHFQPEKNMNDPNGPVYKG 148  
QY 64 AYQLYLHSDONNGPG---GWDHASTTDGVAFTTHGTVMLRPDFF-----VMSGSA 112  
Db 149 WYHLFYQYNEGAIGNKTAHGAHSRDLMLRHLPIAM-----FPDQWYDINGAWSGSA 203  
QY 113 VGTANTAGGAGAVALATQPTDGVKRYQEOYLWSTDGGFTTALPDPIVYN---TDG 169  
Db 204 TV-----LPDGRIVMLYTGSTNA--SVQVQCLAFPSDPS-----DPLLTNWTKEYG 247  
QY 170 RAATTPAEIENAEWFRDPKIH--DTARGEWVCVIG-----RLRYAAFTYSPNLRDWTLLR 223  
Db 248 NPVLYPPPHVGEKDFRPTTAWYDGGDMWRIVIGSKDNRNRRAGMALTYKTKNFHFDELVP 307  
QY 224 NFDYPNHALGGICBDFLFEI-----TADDGTRHWLVA 255  
Db 308 GYLHRVPATGMWECIDILYPVGGARGIDMTAEVAAASNSGGGEVLHVWKESSDDDRH---363

|    |     |   |     |
|----|-----|---|-----|
| QY | 256 | ASMDAYGIGLPMTYAYWTGTWDCGEQHQHADDLITQWLDWGDWDTYAATWPSIDAPETKRLA  | 315 |
| Db | 364 | --DYALG---RYDAATNKWTPLDADADVIGILRYDWG--KEYASKTF--YDPAKKRRVL     | 414 |
| QY | 316 | IAMNNNKYAAROVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALINVTATT       | 375 |
| Db | 415 | WGWGETDSERADV---AKGWASLOSIPRTVVYLDTKTGS--NLIOWPVVEVETLRTNST     | 468 |
| QY | 376 | TLPRDRTVDGSAVLFWN--GRAYEIE-----LDIAWDTATNVGISVGRSPDGRHTNI       | 425 |
| Db | 469 | NLGSIIIVEHGSVFPPLSLHRATQLDIEASFRLLDPLDVAAAKEADGVNGCSTSGGAACRGAL | 528 |
| QY | 426 | GKYG----AD-----LYVDRG--PSDLAGYSLAPYSRAA-----                    | 453 |
| Db | 529 | GPFGLLVLADARRHGGDTEQTAYFYVARGLDGNLRTFHCHDESRSRANDIVKRVVGIN      | 588 |
| QY | 454 | APIDPGARSVHLRILVDVTQTSVEFVNAGHTVLQQQVHFAE---GDTFGISLYTD--GGPAH  | 508 |
| Db | 589 | VPVLDG--EALSVRVLVDHSIVESFAOGGRSVVTSRYPYTEAIVANAGVYLFNNATGARVT   | 647 |
| QY | 509 | FTGIVVREI   | 517 |
| Db | 648 | ATSLVWHEM   | 656 |

```

RESULT 13
US-10-369-493-3411
; Sequence 3411, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3411
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3411

```

[illegible]

```

RESULT 14
US-10-289-757-87
; Sequence 87, Application US/10289757
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Norriss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; FILE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.10610
; CURRENT APPLICATION NUMBER: US/10/289,757
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Lolium perenne
; US-10-289-757-87

```

```

Query Match      10.3%; Score 294; DB 6; Length 668;
Best Local Similarity 24.0%; Pred. No. 7e-14;
Matches 146; Conservative 76; Mismatches 243; Indels 144; Gaps 26;

QY      4  AISRAVLQAGAGALALIFCGAVPPARASAPOSLRAYVHMTPSGWLCDPQRPVTHG 63
Db      95  AASESGMLGADAGSNAPSNAMLOWQRTG-----FHQPEKNMNDPQPVYK 146

QY      64  AYQLYLHSDQNNGPG--GWDHASTTDGVAFTHHGTWPLRPDPF-----VWSGSA 112
Db      147  WYHLFYQNPBGATWGNKNIANGHAVSRDMLRWRHLPTAM-----FPDQWYDINGAWSGA 201

QY      113  VVGTPANTAGFGAGAVVALATQPTDGVKRYQBYLYSTDGGFTFTALPDPIVN-----TDG 169
Db      202  TV-----LPDGRIVMLYTGSTNA--SVQVQCLAFPSDPS-----DPLLTNWTKEYG 245

QY      170  RAATTPAIEIENAEWRDPKTHW-DTARGEVWCVTG-----RLRYAAFTVSPNLKDWTLRR 223
Db      246  NPVLVPPPHVGEKDFRDPPTAWYDGSQGMWRIVIGSKDNRNAGMALTYKTKNFHDFELVP 305

QY      224  NFDYPNALHGIECPDLFEI-----TADDGTRHWVLA 255
Db      306  GVLHRVPATGMWECIDILYPVGGARGIDMTAEVAASNSSGGGCVLHVWKESDDDRH----- 361

QY      256  ASMDAYIGLPMTYAYTGTWGDGQFHADLLTPQLDQMGWDWYAAVTPSIDAPETKRLA 315

```

```
Db 362 ---DYAALG---RYDAATNKTWPLDADADVGIGLRYDMG-KFYASKTF--YDPAKKRRVL 412
QY 316 IAWNNKXYAARDVPTDASDGYNGQNSIVRELRLAROPGGWYTLTSTPVAALTNVYVATT 375
Db 413 WGVGETDSEADV---AKGWSLOSIPRTVLDTKGS--NLIQWPPVEVETLRTNST 466
QY 376 TLPDRTVDGSAVLPMN-GRAYEIE-----LDIAMDTATNVGISVGRSPDGTTRHNI 425
Db 467 NLASIIVEHGSVFPPLSLHRRATQLDIEASFRLLDPLDVAAAKEADVGYNCSTSGGAAGRGA 526
QY 426 GKYG---AD-----LYVDRG-PSDLACYSLAPYSRAA----- 453
Db 527 GPGLLVLAARRHGGTGTATVFFYVARGLDGNLRTFCHDESRSRANDIVKRVVGN 586
QY 454 APIDPGARSVHLRLVDTQSVFVNAAGHTVLSQOVHFAE---GDTGISLYTD--GGPAH 508
Db 587 VPVLDG-KALSVRLVAHSIVESAFAOGRSVVTSRVYPTETAIYANAGVLYFNNTATGARVP 645
QY 509 FTGIVVREI 517
Db 646 ATSLVHKM 654

RESULT 15
PCT-US03-07858-1977
; Sequence 1977, Application PC/TUS0307858
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: PCT/US03/07858
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1977
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Lolium temulentum
PCT-US03-07858-1977
```

```
Query Match 10.2%; Score 291; DB 1; Length 677;
Best Local Similarity 23.8%; Pred. No. 1.2e-13;
Matches 139; Conservative 66; Mismatches 221; Indels 158; Gaps 25;

QY 40 RAVYHMTPPSGWCLDPPORPVTHGAYQLYLHSDQNGPG-----GWDHASTTGGVAF 92
Db 134 RTGFHFQPEKNWNPNGPVYKGNTHLFY----QNPDAIWNKNIANGHAVSRDLVRW 189
QY 93 THRGTVMLRPDPF-----VWSGSAVVGTAAGFAGAGAYVALATPTDGVRYKEQ 144
Db 190 RHLPIAM-----FPQWYDINGAMSGSATV-----LPDGRIVMLYTGSTNA--SVQVQ 235
QY 145 YLWSTDDGFTTALDDPVIVN---TCGRAATTPAEIENAEWFRDPKIHV-DTARGEWVC 200
Db 236 CLAFPSDPS-----DPLLTNNTKYEGNPNVLPYPPPHIGEKDFRDPPTAWYDGSQGMWRI 288
QY 201 VIGRL--RYAAF---YTSPLNRDWTLLRRNFDYPNHALGGIECPDLFEITADGTRHWLA 255
Db 289 VIGSKDDRRHAGMALTYKTNNFIDFELIPGVLHRVPATGMWECIDILYPVGA----- 338
QY 256 ASMDAYGIGLPMTYAYWTGTWGEOPHA-----DGLTPW----- 290
Db 339 ---ARGIDMTAAVAASNNNGGGEVLHVMMKSSDDDDRDHYYALGRYDAATNKKWTPLDADA 394
QY 291 -----LWNGWDWYAAVTPWPSIDAPETKRLAIAWNNKXYAARDVPTDASDGYNCQNSI 343
```

```
Db 395 DVGIGLRYDMG-KFYASKTF--YDPAKKRRVLWGVGETDSEADV---AKGWSLOST 447
QY 344 VRELRLAROPGGWYTLTSTPVAALTNVYVATTLPDRTVDGSAVLPMN-GRAYEIE----- 398
Db 448 PRVVLDNKTGS--NLIQWPPVEVETLRTNSTNLGSIYVDHGSIFPLSLHRRATQLDIEAS 505
QY 399 -----LDIAMDTATNVGISVGRSPDGTTRHNIKYG-----ADLYVDRGSPDLAGYSIAPY 449
Db 506 FRLDPLDVAAAKEADVGYNCSTSGGAAGRGAALGPFGLLVLDADARRHGGDTE---QTAVY 561
QY 450 SRAAAPIDPGAR-----SVHLRLVDTQSVFV 478
Db 562 FYVARGLDGGLRTHFCHDESRSRANDIVKRVVGNVVPVLDGEALSIRVVDHSIVESFA 621
QY 479 NAGHTVLSQOVHFAE---GDTGISLYTD--GGPAHFTGIVVREI 517
Db 622 QGGRSVVTSRVYPTETAIYANTGVLYFNNTATGARVATSLAVHEM 665
```

Search completed: May 6, 2003, 12:43:28  
Job time : 61.6125 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 02:37:06 ; Search time 5273.5 Seconds  
(without alignments)  
11042.896 Million cell updates/sec

Title: US-09-868-328B-2  
Perfect score: 2001  
Sequence: 1 gcggtgaccccgacttccc.....atcgacagctctctgctgtt 2001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description  |
|------------|-------|---------------|--------|-------|--------------|
| 1          | 1566  | 78.3          | 1566   | 1     | AF181254     |
| 2          | 977   | 48.8          | 1890   | 1     | AF294730     |
| 3          | 945.2 | 47.2          | 1752   | 1     | AB001984     |
| 4          | 931.8 | 46.6          | 1551   | 6     | E22826       |
| 5          | 909.6 | 45.5          | 1452   | 6     | E22825       |
| 6          | 370.4 | 18.5          | 1863   | 1     | AB062062     |
| 7          | 147.6 | 7.4           | 65140  | 6     | AX211705     |
| 8          | 147.6 | 7.4           | 123580 | 1     | AF263912     |
| 9          | 147.6 | 7.4           | 123401 | 6     | AX211739     |
| 10         | 141.8 | 7.1           | 3030   | 8     | VCA429230    |
| 11         | 119.8 | 6.0           | 3849   | 6     | AX058889     |
| 12         | 119.8 | 6.0           | 12744  | 1     | AF204401     |
| 13         | 114.2 | 5.7           | 22449  | 1     | SPSNBCDE     |
| 14         | 114.2 | 5.7           | 22449  | 1     | SPSNBCGEN    |
| 15         | 112   | 5.6           | 77457  | 1     | AF210249     |
| 16         | 111.8 | 5.6           | 2951   | 1     | AF123319     |
| 17         | 111.8 | 5.6           | 43147  | 1     | SC4A10       |
| 18         | 109.2 | 5.5           | 27541  | 6     | AX211706     |
| 19         | 106.8 | 5.3           | 30000  | 6     | AX250262     |
| 20         | 106.8 | 5.3           | 88063  | 1     | BSZ94043     |
| 21         | 106.8 | 5.3           | 209510 | 1     | BSUB0018     |
| 22         | 106.4 | 5.3           | 113193 | 1     | AF357202     |
| 23         | 105.2 | 5.3           | 15424  | 1     | AE007019     |
| 24         | 105.2 | 5.3           | 58930  | 2     | AC098321     |
| 25         | 103.8 | 5.2           | 47852  | 1     | MTV023       |
| 26         | 103.6 | 5.2           | 14860  | 1     | AE007093     |
| 27         | 103.6 | 5.2           | 63033  | 1     | MTV008       |
| 28         | 102.4 | 5.1           | 303091 | 2     | AC084799     |
| 29         | 102.2 | 5.1           | 298166 | 2     | AC087563     |
| 30         | 101.6 | 5.1           | 5994   | 1     | SVSNBDE      |
| 31         | 101.4 | 5.1           | 42655  | 1     | SC7H2        |
| 32         | 101   | 5.0           | 15311  | 1     | AE007164     |
| 33         | 99.8  | 5.0           | 4068   | 1     | ANU12274     |
| 34         | 99.8  | 5.0           | 37245  | 1     | SC5F2A       |
| 35         | 99.8  | 5.0           | 40790  | 1     | MTCY493      |
| 36         | 99.4  | 5.0           | 220469 | 2     | AC074307     |
| 37         | 99    | 4.9           | 15348  | 1     | AE007163     |
| 38         | 98.6  | 4.9           | 67200  | 1     | MTV017       |
| 39         | 98.6  | 4.9           | 77534  | 1     | AF235504     |
| 40         | 98.2  | 4.9           | 110000 | 2     | LMFLCHR32_06 |
| 41         | 97.8  | 4.9           | 34108  | 1     | SCE25        |
| 42         | 97.6  | 4.9           | 150754 | 9     | AC023491     |
| 43         | 97.2  | 4.9           | 10420  | 1     | AE005127     |
| 44         | 97    | 4.8           | 4615   | 1     | STMWHB12X    |
| 45         | 96.6  | 4.8           | 46166  | 1     | SC8F11       |

ALIGNMENTS

RESULT 1  
AF181254  
LOCUS  
DEFINITION  
Arthrobacter ureafaciens levan fructotransferase (lfta) gene,  
complete cds.  
ACCESSION  
AF181254  
VERSION  
AF181254.1  
KEYWORDS  
GI:8163731  
ORGANISM  
Arthrobacter ureafaciens  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Micrococccineae; Micrococccaceae; Arthrobacter.  
REFERENCE  
1 (bases 1 to 1566)  
AUTHORS  
Song,K.B., Rhee,S.K. and Yoo,E.J.  
TITLE  
Nucleotide sequence of levan fructotransferase gene (lfta) from

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Arthrobacter ureafaciens K-2032  
Unpublished  
21 (bases 1 to 1566)  
Song, K.B., Rhee, S.K. and Yoo, E.J.  
Direct Submission  
Submitted (27-AUG-1999) Microbial Metabolic Engineering RU, Korea  
Research Institute of Bioscience and Biotechnology, Taejon 305-600,  
South Korea

FEATURES  
source  
gene  
CDS

1..1566  
/organism="Arthrobacter ureafaciens"  
/strain="K-2032"  
/db\_xref="taxon:37931"  
/country="South Korea"  
1..1566  
/gene="lftA"  
1..1566  
/gene="lftA"  
/function="difructose dianhydride IV (DFA IV) formation  
from levan"  
/note="levan-assimilating enzyme"  
/codon\_start=1  
/transl\_table=11  
/product="levan fructotransferase"  
/protein\_id="AAP73829.1"  
/db\_xref="GI:8163732"  
/translation="MTPAISRRVLQAGAGALALIFGAVPPAARASAPGSLRAVYH  
MPPSWLGDPPQVPTTHGAYOLYLHSDNPPGWDHASTDGVAFTHGTVMLPLR  
PPFVMSGSAVNTANTAGFAGAVVALATOPTDGVRYQEOYLYWSTGGFTFALP  
DPVINTDGRANTPAEIEANEFWRDPKLIHNDTARGEWCYVIGRLRYAFYISPNLRD  
WTLRNFDPNALGGIECPDLFEITADDGTRHWLAASMDAYGILPMTYAYWTGW  
DGEQFHADLLTPQULMDWMDYAAVTPWSIDAPETKRLAIAWMNNWKYAAARDVPTDAS  
DGYNGQNSVRELRLARQPGWYTLISTPVAALTYVTATTLPTDVTGSAVLPNWG  
RAVEIEDLTAVRATNVGISVGRSPDTRHTNIGKYADLYVDRGPSDLAGYSLAPYS  
RAAAPIDPGARSVHLRLILDVTSQVFEVFNAGHTVLSQVHFAEGDTGISLYTDGGPAH  
FTGIVVREIGQAI"

BASE COUNT 268 a 635 c 458 g 205 t

ORIGIN  
Query Match 78.3%; Score 1566; DB 1; Length 1566;  
Best Local Similarity 100.0%; Pred. No. 1.1e-160; Indels 0; Gaps 0;  
Matches 1566; Conservative 0; Mismatches 0;

QY 361 ATGACGCGGCCATCTCACGCGCGCGGTGCTCCAGSGAGCGCGCGGAGCACTCGCC 420  
DB 1 ATGACGCGGCCATCTCACGCGCGCGGTGCTCCAGSGAGCGCGCGGAGCACTCGCC 60  
QY 421 CTGATCTTCGGCGGTGCTGCGCGCTGCGAGCCGGCGGCGTCCGCTCGGCGTCCCGT 480  
DB 61 CTGATCTTCGGCGGTGCTGCGCGCTGCGAGCCGGCGGCGTCCGCTCGGCGTCCCGT 120  
QY 481 GCGGTCTACACATGACGCGCGCGGTGCTGCGAGCCCGCAACCGCGGTCACC 540  
DB 121 GCGGTCTACACATGACGCGCGCGGTGCTGCGAGCCCGCAACCGCGGTCACC 180  
QY 541 ACCACGCGCGCTTACAGCTGTACTACTGCTGCTGCGAGCCCGCAACCGCGGCGGC 600  
DB 181 ACCACGCGCGCTTACAGCTGTACTACTGCTGCTGCGAGCCCGCAACCGCGGCGGC 240  
QY 601 TGGGACACCGGAGCAGCAGCGGTGCGCTTACGACCCAGCAGCAGCGGTGATGCCG 660  
DB 241 TGGGACACCGGAGCAGCAGCGGTGCGCTTACGACCCAGCAGCAGCGGTGATGCCG 300  
QY 661 CTGCGGCGCGCTTACAGCTGTACTACTGCTGCTGCGAGCCCGCAACCGCGGTCACC 720  
DB 301 CTGCGGCGCGCTTACAGCTGTACTACTGCTGCTGCGAGCCCGCAACCGCGGTCACC 360  
QY 721 GGGTTCGGCGCGCGGTGCTGCGCGCTGCGAGCCCGCAACCGCGGTCACCAG 780  
DB 361 GGGTTCGGCGCGCGGTGCTGCGCGCTGCGAGCCCGCAACCGCGGTCACCAG 420  
QY 781 TACCAGGAGCAGTACTCTACTGCTGCGAGCGCGGTTCAGTTACCGCGCTGCC 840  
DB 1501 TACCAGGAGCAGTACTCTACTGCTGCGAGCGCGGTTCAGTTACCGCGCTGCC 1560

|                       |   |  |
|-----------------------|---|--|
| Qy                    | 1921  | ATCTAG 1926  |
|                       |   |  |
| Dd                    | 1561  | ATCTAG 1566  |
|                       |   |  |
| <hr/>                 |   |  |
| RESULT 2              |   |  |
| AF294730              |   |  |
| LOCUS                 | AF294730  | 1890 bp DNA linear BCT 15-NOV-2001                               |
| DEFINITION            | Microbacterium sp. AL-210 levan fructotransferase (lftm) gene,  |  |
|                       | complete cds.   |  |
| ACCESSION             | AF294730  |  |
| VERSION               | AF294730.1  | GI:9885836   |
| KEYWORDS              |   |  |
| SOURCE                |   |  |
| ORGANISM              |   |  |
|                       | Microbacterium sp. AL-210.  |  |
|                       | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;   |  |
|                       | Actinomycetales; Micrococccineae; Microbacteriaceae; Microbacterium.  |  |
| REFERENCE             |   |  |
| AUTHORS               | 1 (bases 1 to 1890)   |  |
| TITLE                 | Cha,J., Park,N.H., Yang,S.J. and Lee,T.H.   |  |
|                       | Molecular and enzymatic characterization of a levan   |  |
|                       | fructotransferase from Microbacterium sp. AL-210  |  |
| JOURNAL               | J. Biotechnol. 91 (1), 49-61 (2001)   |  |
| MEDLINE               | 21413997  |  |
| PUBMED                | 11522362  |  |
| REFERENCE             |   |  |
| AUTHORS               | 2 (bases 1 to 1890)   |  |
| TITLE                 | Yang,S.J. and Cha,J.  |  |
|                       | Direct Submission   |  |
| JOURNAL               | Submitted (09-AUG-2000) Microbiology, Pusan National University, 30   |  |
|                       | Changjeon-dong, Keumjeong-ku, Pusan 609-735, Republic of Korea  |  |
| FEATURES              |   |  |
| source                | Location/Qualifiers   |  |
|                       | 1..1890   |  |
|                       | /organism="Microbacterium sp. AL-210"   |  |
|                       | /strain="AL-210"  |  |
|                       | /db_xref="taxon:135249"   |  |
| gene                  | 181..1773   |  |
|                       | /gene="lftm"  |  |
| CDS                   | 181..1773   |  |
|                       | /gene="lftm"  |  |
|                       | /codon_start=1  |  |
|                       | /transl_table=11  |  |
|                       | /product="levan fructotransferase"  |  |
|                       | /protein_id="AAG01554.1"  |  |
|                       | /db_xref="GI:9885837"   |  |
|                       | /translation="MTESFSRRTMLKGTSVGALLIGSSPTLAQSAGLRAPAAA<br>SGSLRAYHMTSPSCWCDPQRVPTGGAYGLYLSHGONNGPGGDHATTSDGVAFT<br>HGVTVMPLODPFVWGSGAVVDNTAGTGCAGAVIALOPTDGIKKYOQLYLWST<br>GGTYFTALPDPLTLNDGRTATPAETENAEWRDPKHHDALRGWEVCVIGRARVS<br>FYTSNLRDQWTSFNFDYPNHALGGIECPDLPMTAGDGTRHWILAAADAYGVGLPM<br>TYATWTGNVTNFIAGNLVPQWLMDGWMAAVTPKAEAPETRRLAIAMNNWKYA<br>ARDVPTDASDYGNSITRELRLEREGGWYLLSEPVGALGGYYTATTIPIOVVN<br>GSYVLPSWRARAYFIELDISWTDAAENVGLSVRSADGARHTNICKYGSGLYVDRGPSDO<br>AGFALSPEPTRAAAPIDAGVSRVHLRVFDTQSVVEVFVNAGHTVLSQQVHFLPGDGTGIS<br>LYSDGGAATFGSIRIREFTPI" |  |
| BASE COUNT            | 318 a 663 c 610 g 299 t   |  |
| ORIGIN                |   |  |
| <hr/>                 |   |  |
| Query Match           | 48.8%; Score 977; DB 1; Length 1890;  |  |
| Best Local Similarity | 78.3%; Pred. No. 4,5e-97;   |  |
| Matches 1170:         | Conservative 0; Mismatches 324; Indels 0; Gaps 0;   |  |
| <hr/>                 |   |  |
| Qy                    | 436   | GCTGTGCGCCTGCAGCCCGGGCATCCGCTCGCGGGCTTCGCTCGCTGCCGTTACCACATG 495 |
|                       |   |  |
| Dd                    | 283   | GCGCGCCTCGGGGGCCCCGGCGCGCATCCGGATCCCTCGCGCGCGTGTTACCACATG 342    |
|                       |   |  |
| Qy                    | 496   | ACGCCCCCACAGCGCTGGCTCTGCGACCCCAACGCCCGGTACACCCACGGCGGCTAC 555    |
|                       |   |  |
| Dd                    | 343   | ACCCCGCGTCCGCTGGCTCTGCGACCCCGCGCGCCCGTGGCGACAGGCGCGCGTAC 402     |
|                       |   |  |
| Qy                    | 556   | CAGCTGTACTACCTGCACCTCGGACCAACAACGGCCCCGGCGGCTGGGACCAACGCGAGC 615 |
|                       |   |  |
| Dd                    | 403   | CAGCTCTACTACCTCCACTCGGACCAACAACGGCCCCGGTGGATGGGACCAACGCCACC 462  |
|                       |   |  |





```

1170 TCCGGCGCTTTCGAATATGCACTCCAGCACACCTTCGGGACCGCACAGTCAACGG 1229
QY 1512 CAGCGCGCTTCCTGCAATGAAGGAGCGCATACAGAGATCGAGCTCGACATCGCTGGGA 1571
Db 1230 CAGTTTCGTACTTCCTGGAGCGCGCGGTATGAACATGGAACATCGATATTTTCATGGGA 1289
QY 1572 CACCGCAGAAAGTCCGGCATCTCGGTGGCGCGCTCCCGCAGGAAACCGGCAACAGAA 1631
Db 1290 CACGCGAGCAACGTGGAGTCTCGGTGGCGCGCTCGTCCGATGGCAGCGCATACGAA 1349
QY 1632 CATCGCAGTACGAGCAGACCTGTACTCGACGAGAGACCTCGGAGCTCGCGGGTA 1691
Db 1350 CATCGCAAAATCGGTAGCAGAGTTTACGTGATCGCGATCGCGATCTCGGAGCAAGCGTTA 1409
QY 1692 CTCGTGCGCCCTTACTCGCGAGCGCGCGCCCGCATCGACCGCGCGCGCATCGTGCA 1751
Db 1410 TCGGCTGGCAGCCCTACACCGCGCGCGCGCGCCCGCATCGATGAGAGCGCAGATCGTGCA 1469
QY 1752 CTTGCGCATCTCTGCGACACCGAGCGTCTGAGGTCTTCGTAACGCGCGCGCACACCGT 1811
Db 1470 CTTGCGCATCTCTGAGACACCGCAAGTCTTGAGGTGTTTCTGTAATTCGCGGCGCACACGT 1529
QY 1812 GCTCTCCGAGAGTCCACTTCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1871
Db 1530 GGTTCGCGAGAGGTGCACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1589
QY 1872 CCGCGCGCGCACACTTCACCGCGCATCTGCTCGCGCGAGATTCGCCAGCGATCTAGCGAT 1931
Db 1590 CGGTGCGCGCACTTCACCGCGCATCACCATCCGCGAGTTCGCGGAGCGCGCGCGCG 1649
QY 1932 GCACACACACCGCTCACCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1975
Db 1650 CGTCCACCGCTGGAAGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1693

RESULT 4
E22826 Levan fructotransferase gene. 1551 bp DNA linear PAT 18-JUN-2001
LOCUS E22826
DEFINITION Levan fructotransferase gene.
ACCESSION E22826
VERSION E22826.1 GI:13024125
KEYWORDS JP 1999069978-A/2.
SOURCE Arthrobacter nicotinovorans.
ORGANISM Arthrobacter nicotinovorans
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococciaceae; Micrococciaceae; Arthrobacter.
REFERENCE 1 (bases 1 to 1551)
AUTHORS Masakazu, S., Atsushi, Y. and Fusao, T.
TITLE Levan fructotransferase gene
JOURNAL Patent: JP 1999069978-A 2 16-MAR-1999;
NIPPON OIL CO LTD
COMMENT OS Arthrobacter nicotinovorans
PN JP 1999069978-A/2
PD 16-MAR-1999
PF 28-AUG-1999 JP 1997232421
PR MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TONITA
PC C12N15/09, C07K14/195, C12N1/21, C12N9/10//A61K38/45, (C12N15/09,
PC C12R1:06),
PC (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,
PC (C12N15/00, C12R1:06)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source i. .1551
FT Location/Qualifiers
FT 1. 1551
FT /organism="Arthrobacter nicotinovorans"
FT /db_xref="taxon:29320"
BASE COUNT 315 a 524 c 443 g 269 t
ORIGIN

```

```

Query Match 46.6%; Score 931.8; DB 6; Length 1551;
Best Local Similarity 75.8%; Pred. No. 3.6e-92;
Matches 1184; Conservative 0; Mismatches 367; Indels 12; Gaps 2;
QY 361 ATGACGCGCGCATCTACGCGCGCGCGTCTCCAGGAGCGCGCGCGCGAGCATCTCGC 420
Db 1 ATGACGCTATGACATCTCTCGCGCATCTCGCGCGCATCTCGCGCGCATCTCGCGCGCAT 60
QY 421 CTGATCTTCGCGCGGTCTCTCGCGCGCATCTCGCGCGCGCATCTCGCGCGCGCATCTCG 480
Db 61 CTTTTCATGAGCAATGCGCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 114
QY 481 GCGCGTCTACCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 115 GCGATCTACCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 174
QY 541 ACCAGCGCGCGCTACAGCTGTACTACTGTACTCTCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 175 ACAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 234
QY 601 TGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 235 TGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 294
QY 661 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 295 ATGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
QY 721 GGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 355 GCGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 408
QY 781 TACGAGGAGCGCTACTCTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Db 409 TTCAGGAACAGTACCTTTACTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468
QY 841 GACCCCGCTCATCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db 469 GACCCCGCTCATCTTGAACACTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528
QY 901 GCGAGTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Db 529 GCAGAATGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 588
QY 961 GTATCGGAGCGCTGCGGTACCGCGCGGTCTACACTCGCGCGCGCGCGCGCGCGCGCG 1020
Db 589 GTATCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
QY 1021 CTTTCGCGCGCAACTTCGACTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db 649 TGAAGTCCCACTTCGACTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
QY 1081 TTCGAGATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db 709 TTCGAAATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768
QY 1141 TAGCGCATCGCGCTCCCGCATGAGCTACGCGTGTGACAGCGCGCGCGCGCGCGCGCG 1200
Db 769 TACAGCATCGCGTTCGCCCATGACCTTTCGCTTACTGACAGGTTCATGGAACCGCG 828
QY 1201 TTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
Db 829 TTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 888
QY 1261 GTCACTGTGCCATCGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Db 889 GTGACTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
QY 1321 AACTGGAAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 949 AACTGGAATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1008

```

QY 1381 AACTCGATCGTCGGGAGCTGGCGTCCGCCGACAGCGCTGGCGGCTGGTACACCCCTCCTG 1440  
Db 1009 AATTCCATACGCGGAGCTCAGCTCGAGCGCAATCGCGCGCTGGTACACCTTGTCT 1068  
QY 1441 AGCACCCCGTGGCGAGCGCTGACGAATAGCTCACCGCCACCAACACACTCCCGGACCGG 1500  
Db 1069 AGCAGCGCGTTCGGCGCTTTCGAACTATGCCACCTCCAGCACACCCCTTCGGACCGC 1128  
QY 1501 ACCGTGACGGCAGCGCCGCTCTGCCATGGAACGGACGGCATPACGAGATCGAGCTCGAC 1560  
Db 1129 ACAGTCAACGGCAGTTTCGTACTTCGTGAGCGCGCGGCGTATGAATGGAACCTCAT 1188  
QY 1561 ATCCCTGGGACACACCGGACGAACTCGGATCTCGTGGCGCGCTCCCGGACGAACC 1620  
Db 1189 ATTTCATGGGACACGCGACGAACTCGGAGTCTCGGTGGCGGCTCGTCCGATGGCAGC 1248  
QY 1621 CGGCACAGCAATGCGCAAGTAGCGAGGAGACCTCTAGTTCGACGAGGAGACCTCCGAC 1680  
Db 1249 CGGCATAGGACATCGGCAATACGTTGAGGATTCGATCGGATCCTCGGAG 1308  
QY 1681 CTCGCCGGTACTCGCCGCCCTACTCGGAGCGCGCCGCCCTACGACCCCGCGCC 1740  
Db 1309 CAAAGCGGTATCGCTGGCACCTACACCGCGCGCGCGCCATCGATGGAACGCC 1368  
QY 1741 CGATCGTGCACCTCGCATCTCTGTCGACACCGAGCGTTCGAGTCTTCGTCACGCC 1800  
Db 1369 AGATCGTGCACCTCGCATCTTTGTAGACACCAAGTGTTCGAGTGTTCGTAATTC 1428  
QY 1801 GGCACACCGTCTCTCCAGCAGGTCCTACTCGCGAGCGCGCGCGCGCGCGCGCTC 1860  
Db 1429 GGCACACCGTGTTCGACGAGTGCATCTCGCGCGCGCGCGCGCGCGCGCGCGCTC 1488  
QY 1861 TACACGACGGCGGCGCGCGCACATTCACCGGATCGTCTCGCGAGATGTCGCGAGCG 1920  
Db 1489 TATCGGAGCGCGCTCGCGCAACTTCACCGGATCACCATCGCGAGTTCGCGAACCC 1548  
QY 1921 ATC 1923  
Db 1549 ATC 1551

RESULT 5  
E22825  
LOCUS                    1452 bp   DNA    linear   PAT 18-JUN-2001  
DEFINITION            Levan fructotransferase gene.  
ACCESSION            E22825  
VERSION               E22825.1 GI:13024124  
KEYWORDS              JP 1999069978-A/1.  
SOURCE                Arthrobacter nicotinovorans.  
ORGANISM              Arthrobacter nicotinovorans  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micrococciaceae; Micrococaceae; Arthrobacter.  
REFERENCE  
AUTHORS               Masakazu, S., Atsushi, Y. and Fusao, T.  
TITLE                  Levan fructotransferase gene  
JOURNAL                Patent: JP 1999069978-A 1 16-MAR-1999;  
COMMENT                NIPPON OIL CO LTD  
OS                    Arthrobacter nicotinovorans  
PN                    JP 1999069978-A/1  
PD                    16-MAR-1999  
PF                    28-AUG-1997 JP 1997232421  
PR  
PI                    MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TOMITA  
PC                    C12N15/09, C07K14/195, C12N1/21, C12N9/10//A61K38/45, (C12N15/09,  
PC                    C12N1:06),  
PC                    (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,  
PC                    (C12N15/00, C12R1:06)  
CC                    Strandedness: Double;  
CC                    Topology: Linear;  
FH                    Key            Location/Qualifiers  
FT                    source        1. .1452  
FT                    /organism= 'Arthrobacter nicotinovorans',  
FEATURES              Location/Qualifiers

source            1. .1452  
                  /organism= 'Arthrobacter nicotinovorans'  
                  /db\_xref= 'taxon:29320'  
BASE COUNT       299 a   491 c   416 g   246 t  
ORIGIN  
Query Match       45.5%; Score 909.6; DB 6; Length 1452;  
Best Local Similarity 77.0%; Pred. No. 9.3e-90;  
Matches 1123; Conservative 0; Mismatches 329; Indels 6; Gaps 1;  
QY 466 CGGGGCTCGCTCGTCCGCTTACCACATGACGCCGCCCGCCAGCGGCTGGCTCTCGGACCCC 525  
Db 1 CAGGATTCCTTCGGGGATCTACACATGACGCCGCCCGCCCTCGGGCTGGCTATGTGATCG 60  
QY 526 CAAACGCCGCTACACCCACCGCGCTACCACTGTACTACTGTACTGTACCTCCGACCGAAG 585  
Db 61 CAGCGACCGCTACATACAAAGGCGCTACCAGCTCTACTACCTCCACTCCGCGCAGAAC 120  
QY 586 AAGGCGCGGCGGTGGGACACGCGAGACGACGAGCGGCTCGCTTTCAGCGACGAC 645  
Db 121 AAGCGACCGGCGGATGGGACACGCGACGCGGCGGAGGTGTCTTACACCCACCAT 180  
QY 646 GGCACCGTATGCTCGCTCGGCGCGGCTTCCCGCTGTGTTCGCGGTCCGCGGTCTCGGC 705  
Db 181 GGAGTGTGATCCAAATGCAACCGGACTTCCCGCTGTGTTCGCGGTCCGCGGTAGTGGAC 240  
QY 706 ACCGCAACACGCGAGGTTTCGGCGCGCGGCTCGCGCTCGCGACCCAGCGGAC 765  
Db 241 ACCGCAACACGCGCGCTTCGGCGCGCGGCTCATCGCGCTCGCGACCCAGCGGAC 300  
QY 766 GAGCGGTCCGAACTACAGGAGAGTACTCTACTGTGACCGACGCGGTTTCAAG 825  
Db 301 GAGC-----GAAATTCAGGAACTACTCTTACTGTCTCCACGATGCGCTACTCC 354  
QY 826 TTCACCGCTTCGCCACCGCGTCTGTCACACGAGGTGCGCGCGCGCACCGCGCC 885  
Db 355 TTCACCGCTTCGCCACCGCGTCTGTCACACGAGTGTGAACTGTGAGGAGGCGCGCACCG 414  
QY 886 GCGGATCGAAGCGCGGAGTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
Db 415 GCGGAGTGGAGAACGAGAAATGTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
QY 946 GGAGATGGTCTGCGTATCGGACGACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1005  
Db 475 AAGAGTGGTCTGTGTCATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534  
QY 1006 CTGCGCGACTGGACACTTCGCCCGCAACTTCGACTACCCGAAACACCGCCCTCGCGCGCATC 1065  
Db 535 CTGCGGATTTGCAATGGAATCCAACTTCGACTACCCCAACCGCGCGCGCGCGCG 594  
QY 1066 GAGTGGCGCGCTGTGTCGAGATCACCGAGAGCGGAGCGGACGCGCGCTGGGTGCTCGCC 1125  
Db 595 GAATGCGCGGATCTGTTCGAAATGACCGAGGAGAGGAAACCGCGCGCGCGCGCG 654  
QY 1126 GCCAGATGAGCGCTTACGCGCTCGCGCTCCCATACGCTACGCTACTGTGAGCAGCG 1185  
Db 655 GGAGCATGAGCGCTTACGAGCTCGCGCTTGGCCATGACCTTTCGCTTACGAGAGTTCA 714  
QY 1186 TGGGACGCGGAGCTTCCACGCGCGGACGACTCACCCCGCAATTCGCTCGACTGGGCTGG 1245  
Db 715 TGGAGCGGACAGCATTCATCGCGGACAACTTCACACGAGTGGCTTGCATGGGATGG 774  
QY 1246 GACTGTAGCGCGCGCTTACCTGGCCATCGATGAGCGCGCGGAGACCAAGCGCTCGCC 1305  
Db 775 GACTGTAGCGCGCGCTGACCTGGCGCGCGCTGGAAAGCACTGAGACCAAGCGCTTGGC 834  
QY 1306 ATCGCTGATGAACACTTGAAGTACGCGCGGAGTACCGCGCGGAGTCCCGCGCGCGCG 1365  
Db 835 ACAGCGTGTGATGAACAACTGGAAATATGCGCGCGCGCAACGTCGCCACGAGCGCTCCGAT 894  
QY 1366 GGTCTAACCGGAGAACTTCGATCGCTCGCGAGCTCGCGCTCGCGCGAGCGCTCGCGCG 1425  
Db 895 GGTCTAACCGGAGAAATTCGATCGCGCGGAGCTCAGGCTCGAGCGCGCAATCGCGCGCG 954



Db 682 GACGACGAGCGCCAGCGCTGGGTGCGCCCTCATCGCCGAGCGGACCGGTTTCGTCTTAC 741  
QY 994 ACCTCGCCGAACTTGGCGACTGGACACTTCGCGCAACTTCGACTACCCGAAACACGCGC 1053  
Db 742 ACCTCGCCGAACTTGCACCGCTGGAC-----CGGACGGCGGAGTACGTACAGCGCGC 795  
QY 1054 CTCGGGGCATYAGTGGCCCCCGACCTTTCGAGATCACCGCAGACGAGCGACGCCAC 1113  
Db 796 ATCGGCACGATGAGTGGCCGCGACCTTTCGCGCTTCGGGCGCAGCAGCGACGCCAC 855  
QY 1114 TGGGTGCTCCCGCCAGCATGGAGCTAGCGCATCGGCTTCCCTCCCATGAGCTAGCGCTAC 1173  
Db 856 TGGGTGATGGGCTCAGCGGAAAGGGTACGCGACGAACCCCGGACCTACGCGGTAC 915  
QY 1174 TGGACAGGCACTTGGAGCGCGAGCACTTCCACGCGCGAGCACTCAACCCGCAATGGCTC 1233  
Db 916 TGGACGGCTCGTTGACGCGGTGACCTTCTGTCGCCGAGCTGCTGCGCGCAGTGGCTC 975  
QY 1234 GACTGGGGCTGGAGTGGTACGGCGGCTGACCTTGCATCGACGCGCGCGAGACC 1293  
Db 976 GACACAGGTTGACTGGTACGGCGGCTGACCTGGGAGGACCCCGCGCAGCTCGAG 1035  
QY 1294 AAGCGCTCGCCATCGGTGATGAACACTGGAAGTACCGCGCAGCGAGTCCGCCAC 1353  
Db 1036 CGCGGCTCGGATCGGCTGGATGAACACTGGGACTA---CGGACGGCGCGCGAC 1092  
QY 1354 GAGCGATCCGACGGCTACAAAGCGGAGAACTGATCGTCCGCGAGCTGCGGCTCGCCGA 1413  
Db 1093 TGGCGCGGAGCGGATTCACCGCGCAGGACTGATCAGCGCGGAGATCCGCGCTCGCCG 1151  
QY 1414 CAGCTGGCGGCTGGTACACCTCTTGAGACACCCCGTGGCAGCGGTGAGCAACTAGTTC 1473  
Db 1152 --TGGCGTCCGGCTACAGCTTCTTCCGCTCCCGTCCGCTGACGCGCTTCCAGGACCGG 1209  
QY 1474 ACCGCGACCAACACTCCCGCAGCGAGCTGATCGTCCGCGAGCGCGCTCTCCATGGAAC 1533  
Db 1210 ACCGACAGAGCGGCTTGGCGAGCTGCGGCTGACGCGGTTCACGAGCTGCGCTACCGC 1269  
QY 1534 GGAGCGCATACGAGATCGAGCTCGACATCGCTGGGACACCGCGAGCGGCTCGGCATC 1593  
Db 1270 GCGACGCTACGAGATCAGACGCGCATCTCGCGCAGGACCGGACACGCTCGGCTTC 1329  
QY 1594 TCGGTGGCGGCTCCCGCGAGCAACCGCGCACACGAACTCGGCAAGTACGAGGACGAC 1653  
Db 1330 CAGCTGGCGGCTTCCGCGAGCGGACGCGGACGCGGCGGCTTCAACCGCGACTAC 1389  
QY 1654 CTGTAGCTGACGAGGACCTCCGACCTCGCGGCTACTCTGCTCGCCCTTACTCGCGA 1713  
Db 1390 GCCTACTCAACCGGGGACGACCGGCGGCGGCTGCTGGAAGG-----TCGAGA 1441  
QY 1714 GCGCGCGCCCATGACACCGCGCGGCTGATCGTGCATCGTGCATCTCTGCTCGGACAC 1773  
Db 1442 GCGCGACGCGCTTGA---CGAGCGGACACCGTGGAGCTCGGATCTCTGCTGACCGC 1497  
QY 1774 CAGAGCTGAGGTCTTGTACAGCGCGGCGCACACCGTGTCTCCAGCAGGTTCACATTC 1833  
Db 1498 ACAGCATCAGGTCTTGTGCGGACGCGGCGGTACACGACTCTGAGCGAGGTGTTTCGG 1557  
QY 1834 GCGAGGGGACACGGGAATCTCGCTTACACGAGCGGCGGCGCGCACACTTTCACCGGC 1893  
Db 1558 CCGTCGGGGCAACAGGCTTCGCGCTTACACCTGGGCGGCGCTCGGTGTTTCGCGAT 1617  
QY 1894 ATCGTGTGCGGAGATGGCGAGCGATCTAGGCGATGACACACGACCGCTTCACCGGAA 1953  
Db 1618 CTCGGATCACCAGTTCGAGACCTTCGCGAGCTCCGCGGCGCTTCTCGGACTTCG 1677  
QY 1954 GCGCGCGCGGAGACGAGCG 1975  
Db 1678 AGGCGACACGTGGGCGGCGG 1699

RESULT 7  
AX211705

LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 1 from Patent WO0159126.  
ACCESSION AX211705  
VERSION AX211705.1 GI:15523937  
KEYWORDS Streptomyces noursei.  
SOURCE Streptomyces noursei.  
ORGANISM Streptomyces noursei  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 65140)  
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,  
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and  
Gulliksen,O.M.  
TITLE Gene cluster encoding a nystatin polyketide synthase and its  
manipulation and utility  
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;  
Norwegian Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN  
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOESKOLE (NO) ;  
ALPHARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich  
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;  
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein  
(NO)  
FEATURES  
source location/Qualifiers  
1..65140  
/organism="Streptomyces noursei"  
/db\_xref="taxon:1971"  
/note="ATCC 11455"  
BASE COUNT 8270 a 25171 c 22273 g 9426 t  
ORIGIN

Query Match 7.4%; Score 147.6; DB 6; Length 65140;  
Best Local Similarity 45.4%; Pred. No. 5.7e-08;  
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;  
QY 8 ACCCGGACTTCCCTCGAGACACCGTCCCTACCGCGGACCGCGCGCGGCGGACTGC 67  
Db 19803 ACGGCGGAGACACGACCGCGCGCTTCGGACTCGACCGCGGCTTCGGAGCGCGCAC 19862  
QY 68 TCTCAGCTAGACGGCGGCTCTCTCGAGGTCTTCGTGGGGACGCTGAGCGACTGCGT 127  
Db 19863 AACACGCGCGGCTACGCGGACCTCGGCGCATACGCGCGGCGCTTCGCGCT 19922  
QY 128 CGAACCTGCTCTTCGCGGGCGGCGGTGTACCGGGAGCTTCGAGACGGACGCGCAG 187  
Db 19923 GGAAGCGCTCTCGTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19982  
QY 188 GAACCTGTGACGTGACCGGCGTACGTGAGGCGCGGCGGCGGCGGCGGCGGCGG 247  
Db 19983 CCGCGGAGGACACCGGTACCATCGCGTCTACGCGCGGCGGCGGCGGCGGCGG 20042  
QY 248 CTGCGCGGCTTCGCGCTGACGAGCGCTCCACCGCGGAGAGTCTCTCTTACCGTGC 307  
Db 20043 TCGACTCTCTGGTCTCCCGGAGTCCCGCGGAGCGACCGCGCGCGCGGCGGCGG 20102  
QY 308 CGAACGAGGTGACCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 367  
Db 20103 ACCGCGACTCTCTTTCACGTGAGTGGACCGCGGCGGCGGCGGCGGCGGCGGCG 20162  
QY 368 CGGCGCATCTCACGCGCGGCGGCTCTCCAGGAGCGCGGCGGCGGCGGCGGCGGCG 427  
Db 20163 CGGCGCA-----CCGTCGCGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 20213  
QY 428 TGGGCGGTGTGTGCGCGCTTCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487  
Db 20214 TCGCGG-----CCACCGGATCCGGACACCGCGGCGGCGGCGGCGGCGGCGG 20264  
QY 488 ACCACATGAGCGCGCGCGGCGGCTGTGTGACCGCGGCGGCGGCGGCGGCGGCGGCG 547  
Db 20265 CCGACCGCGAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20321  
QY 548 GCGCGCTACGAGTGTACTACCTGCGACTCGGACGACGCGGCGGCGGCGGCGGCGG 607  
Db 20322 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20381



```

/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
/translation="MASPDLEERTAPRRRLRGLLRPHRRSVALVSMGVGGIVL
NAFPLLLGRVTDLLADGVLGGVPGAPGIDFAAIGRLLLVLLVYVASLFLAQR
LVASAVRTIHELRLDAGREKLTPLRHFDFRAGELLKRTINDINLOOTLOOTLAE
LITSFSLTMLVLMVISPVLAVMLLSVPVSAITAARISKRAOPHYAAOSANGTL
NARVEVCTGHALIGEDRRRAAEFNDVNDVRAAKAQAQFASGAPVMMFVANL
GVVLVACIAMKVINGLTLGDVQAFILYARQFISOPITELIASVAGLQSGIASAORVF
TLLDAPEQADPLRCPCTPARAEGRVEFTDVSFRYSDPTLJENLSLTVPEGTSOAIVG
PTGACKTTLGNLLMRYPEDSGRILLDGDGTATMTDRDRLSRFGLVLOQTWLFEGGTIA
ENIYAGACACADIEEARAATCADRFITRLPGQVITLDDDESGIVSAGEKOLLTVAR
AFLARPAVLVDEATSSVDYTRTEVLIQRAMLSLRAGTSTFIAHRLSTTRDADLIYVM
DAGRIVEOCTHQDLCAQGLYARLHAARTHTPTAGAAAQ"
complement(2662..4416)
/gene="nysg"
CDS
complement(2662..4416)
/gene="nysg"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AAF71764.1"
/db_xref="GI:8050838"
/translation="MLLRLLRAQLRPYAWATAALVALQVLQILGTLTLLPTLGAALIDQ
GVVRDGGRTITELGVVMVVALVOIAAALGAALAAATATAMGRDLSALFRILDF
AREIGRFGTSLTTSVNDVQVONLAQTGFGIVCAPLCLGSLVLLALRQYPLALL
LVALVAVACGLLARLMTLYARMLQTLDRGLRLREAITGVRVRSFVRDHERA
RQAQNDLVVSRVRGLTIATPVLVLLMNGFTVALLWTSGRHDAGRMPIGSLA
LISLILMSVLMIAFVLPVPRARVCAGIAEVLDTGSSVAPPAQPVRPGPACRI
ELCAAGYRPGAEPEVLDVLTVEPGERIAVLGSGTKTLLNLVRLADATGAV
RYGTDVRELTAATAAAVGFVPPRPLFSGTVASNLRRGRDPDATDEELWEALRVAQA"
ADFVARMPPGLDAELTGGGNNVSGGQRLSLARALLRRPEIYLFDDCFESALDQNTDA
ALRTALPYTAGATVITVAQRISAGRDADRIIVLDGRVVAAGYTHDVLRLTPTREI
ALSQUTEEEAHGLAGR"
4714..5748
/gene="nysDIII"
4714..5748
/gene="nysDIII"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dGDP-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysDIII"
/protein_id="AAF71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGQDPSYLAHLLSQGVQVWGLIRGOANPKRSRV
SLRASELDFIDGDLMDQGSLSVSAVDTVQPDVYNLGAISFVPMWQQAELTEVNGMG
VLRMLEAIRMVSGLSTRVSPRGQIRFYQASSEMFGKAETPQRETTLHPHRSYG
AAKAYGHVITRNYRESFGMYAVSGMLNHESPRGQEFVTRKISLAVARIKQGLQDL
ALGNLDVARDWCYAGDYVRAMHLMQDAGDDYVLTGTOMHSVRDVARIPAEHVGILNW
EDYVVDPLVPRPAEVEVLCADSAQDLRLGKPKVDVPTLNRMMVDSDLQAGVSRNO
YGDVLLAANW"
5930..34363
/gene="nysi"
5930..34363
/gene="nysi"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="Nysi"
/protein_id="AAF71766.1"
/db_xref="GI:8050840"
/translation="MDNECKLRDYLKLATADLRRTRRRVHKLESAQEPVAIIQWTCR
YPGVGRSPEDLRWMEAGEHGVTPEPTDRGMDLEALAAAPTASGGLFHDAPDFDFF
GISPREAVAMPQQRVLESWEAEPFRAGIDPYSKSLSVGIFIGAMAQDYRVGPDG
AEGFOLNTGTSGLSRIYSYFTGTVGPAVTVDTSKSSLVAVHLIATQALRAGECTLAL
AGGVTIMSGPGTFIEMGRQGLSADGRCRSGDGTADGTGMAEGVGLVLERLSDAVRN

```

```

GHEILAVRGCTAVNODGASNGLTAPNGPSQOQVIQOALVHARLAAGDIDVVEAHGTGT
TLGDPVEAQLLIATYQGNRPADRPDLILSGVSKNSHTQAAAGVAGVVKWMAMRRHGT
PRTLHAEPTTHVDMSQAVRLLTDTDPWATCAPRRAAVSFGISGTNAHTLIEQAP
EPQPEDAATQDDAAGSTPATVPVGVVLLSGRTPDALRQDADPPAAAAPWITGTTRAP
DULDLAHSATTRAGFEHRAVLATDHPALTDLGTLTALADDDPAAAAPWITGTTRAP
RLAYLPTGGOAORLGCAGRELAARFAPAFALDAADAFTHLDRPLREVILWGTDAALL
RATVQAQPALFAVEVALYRLIESFVGRPHLAGHSVEIVAHLACVLSLADATLVA
AKRMLQALPDGGAMIQAQSEADAPLILAGHEDVIAIAVNGPSVLSUAGAEATVTA
LAEQAADGKRTLRRLVSHAFSPIMEPLDAFRAVVEDTLQPLLPPVSNLTKGPA
TVAQILTADYVWVHVRHAFADGIDWLARHDTTAPLELPGDGVLSAMAQDCIDADA
TLPVLRAGRPEREHTLTLAGLHLHVGATLDWTCGCTAGTACTRTDLPITYAFORRY
WPKALQSCADLRSVGLCAAHUPLLSAASLADAGCTLLTGLRSLRQTHPLADPVRV
TYLLPGTAFLELAVRAGDEVGCDRVEELTAAPLLLPEQGGVQVQLMIGNPNVSGRRT
VNWHPARDPDDTPTAHATGVLTTADASRLQRPASSSEQGTPLAGDPHALDAOAMP
AGAEPFLDGHYDLRLADGGFGYGVFQGLRAAWRGDDVGYIAEVELPEAGRSDAEFL
HPALLDAALHAAFPFTGLGRGRGLPFSWEGVSLHAGGATLRLVRLTPVADALATV
ADGTGAPVLSVDSLVRVATQOGLDTAAVARDALFLDWTVPQPTATDPGVALIGA
DFPGLLTHAGFADAPAYPDLAALAAGDPVTIVVLSLACTGDDAADPARSAHRCAR
ALAAVQTLWIDHHERFAARLVFVTRGATVGRDVAANAAMVGLVSAQSENPCCFALVDL
DPDGAAGAAALVAALVSGEPQLAVRGDVLRLVRLRRLRTEVAGADGTGDDVGGSG
VSFGVAGLVTTGGTGGGLGAVLARHLVAVVHAAGVLDLDDMGALTCERLSAVLRKVD
GARVGVACDVTDRAAVVELGVGHASVAVHAAGVLDLDDMGALTCERLSAVLRKVD
AVMHLHEATRGDLDLDAFVFSVLGAGVSGPOANYAANAFILDALMTRRAREGLPGLS
LANGPWEQSGGHTGLTLDVDAERLARAGVPLSVAQGLALFDAAVAGTDTATCPVRLD
LPVLRAGEVPPLLSLIRVARRAAVAGSAGTANLAQRLRELDGDEMVDLULVRG
QVALVHLHATGGDVADGAFRDFDLSLTVELNRRLNTVTLRLPATLIFYDPTVVRH
LATYVLDELGTDAEVATQPAAVADDPIVIVGMACRYPGVSGSPEDLWRLVTEGT
DVGFPTRNRGDWESLYHPDPHPTGTSYTRSGGLHEAGDFDGFSGMSPREALATD
SOORLLLESSWEAIEFRAGIDPVLSRGTGFGVAGVMSYDYSAMLASPEFEGQSGSS
PSLASRVAYTLGLSGPVTVDTCSSSLVAMHMAQALRSCEGLLAGGVTVMSPT
AVFDFQRGLSPDGRKAFADAADGVSGSEGVLVLERSDADVNRNGHELIAVVRG
SAVNDGASNGLTAPNGPSQOQVIQOALVHARLAAGDIDVVEAHGTGTGDDP IEQAQ
LIATYGRDREPERLLLSVNLGTLTAAGAGVAGVVKMLHMRHGVCPKTLTAADAPS
SHVDSNAGEPRLLSQAAMPETGRVRRAGVSGISGTNVHVIVEQAPCAALAAAGA
ARTPGAVPLLSGRSALRGOAARLHGLQAPDAELVDVALSLATTRSREORAA
VYVARDQLIASIGALADRPVAVVEGEAAGRTAVLFTCGGSGORAAHRELHEVQ
PEFAADAVCAVDFPLDRPREVVEEDGSDERALLDETGTQTPALFAVEALFRL
VESWGRPFDVAGISGETAAHAGVLTLEDCRLVRAARATLMQALPTGGAMIAIQA
TEDEIAHLDDTVAIAAVNGPSQVVISGDEAAETIAATFAERGKRLRLVSHAFHS
PRMDGLDAFRIVAEGLTVRAPRIPLVSLDTRADDADAEVTAEYVVRHVRVREAFAD
CVTRLRDAGATTFLESGDGLTMAEDTLGDHDAELVPLMLRAGREBELAAATLAR
LOVRGVDVMAAYLAGTCARTDLPYAFQAHAYYMPOLPTPAALAAADPAQOOLWAA
VERGDARELADILGIGEOQLTPLDSILLFALTWSRNGQEKHLTLTRYRVENTRISKP
TAPVLDGTLVLSADATADQPALDGLADASHGARGVRRLDLDSCADRAVLAERL
ARTADVDAATQVLSVPLDERDADDCPLTGLTALTVALQADTLAGOGRWLTATRG
AVSTNPADPVTHPVQAAAGLGRGVLEHPRLWGLVLDLTVFDERAGORLAGILAVK
DAPGCEOVALTATGVSGRRLVRHTVEALPTAAETATGTLITGCTGLGAEARWL
APAGAQLVLTSSRGPDPAGAAELRAELEGGYSPSVVACDVADRDALAAVLTALPEE
LPLTGVVHTAGVGHYGLDPLTSTAEFAGLTAAKLAGAHLADLADRELDFVLEFSI
AGVSGSQSAYGAANAYLDLALHRRARGLAATSVAMPAGAEAGVMAADDVASETLRR
QGLGLDPAAMTELRRVAVRODVTTVADVDMQRYAPLFTSARPSALITAGLPEVRAL
ADGTEQDATGASEVTVRVALABPEQLRLTDLVLTESATVLGHSSADAVEGRAE
RDVETSUATVELRKLGAATGLSPVMVFDYPTPLAQLAEALICGLVAVAGPV
ATGCGADDEPIATIGNACRPGCVSSPEOLMDIVASGDAISEEPVNRGWOTGHLDPD
PBRPGTITSTGCGFLHEADEPTEFGISPREALVMDPOORLLLETWRESFERAGIRP
EFLRSTLTGTFVSGSYQTEYGAGDGTGEGHMTVSGPSVLSGRLLTFTVLEGPAGTVD
TACSSSLVALHACVSVLRNENSLAVAGGATIMTTPNPFIAFSRORALAKDGRCKAFS
VTRQALANARLAPGDIDALEHGTGTPLGDPTEAQLFATYGRDRDPSALLGVSYS
NIGHTQSAAGTASVTKMYMALRHSLEPPTLHADAPSSHVDWSAGTVRLTQOARWPE
GRPRAAVSSGISGCTNAHVLEQAPVADTPAERPAPVAPVPIAGVVPVWYVARSAA
ALURGOERLLHAETVGTALPAAGPLDGLSLVSARARFEHRAVVPVVPAGTDPDLAAL
AVATDGPSPVAVGADVDEGRTVFFVPGQSQWGMGSQLDESFAFARIACAAAL
AEFTDWSLVLDLVRGVGAPSLERVDVQFASVAVMSLAALWRSGLVPLDVAVHSHQ
ETAAAVSVGSLSRDGRVARSQAIGRALAGRGMMMSVSLSDVLEPLRVLEFEGVR
SVAAVNGPSVSVVAGEPEALDHALHARTADDIRARRIADVASHSHVEDLHEELLEV
LAELAPRTSEVPEFTVTGDMARTMDAGYHFRNLGRVREADADVLLAEYRAFY
EYSSHPVLSMAVQEAIDEACGPAVAGTLRDQCGTDRFLLSAAEFVVRGVVDWAGL
FEGDSVAAIDLPITYAFQHEHLWAPPAPEAVAAADDDAFTWAVEDGVSAULTAALG
TDTGSVAAFLPALTSWRRLRDRSTVDAMRYAVANKPLGGLTLPHSGLTGLWLTADG
PCTSLVLTALTYGAERVRLVDECEVRLRAGAEADVGTIVSVLAQAERTDVA
PCTSLVLTALTYGAERVRLVDECEVRLRAGAEADVGTIVSVLAQAERTDVA
AALEHPQRWGGTLDLPAALDARAQRLAAVLSGALGSDDLALRFSGVFTRIVAREA

```



|            |       |  |                                   |
|------------|-------|--|-----------------------------------|
| QY         | 10128 | GCACATCTTCGACTATCCGGAACCAACGACCCCTTCGGCGGCACTGAGTGTCCCGGCGGCTCAGCGGCGCAACTCACGGGCTCGCGCGCCA--CGTCC | 807111                            |
| Db         | 80652 | CGGCGCGGCGAGCGGGCGCGACACACTGACGGGCGCAACTCACGGGCTCGCGCGCCA--CGTCC                                   | 807111                            |
| QY         | 10088 | TCACGCGAGACGAGCGGACACGCCACTCGGTGCTCTCGCGCGCAGCATGACAGCGCTACAGGCA                                   | 1147                              |
| Db         | 80712 | ACATCGCGCTGCGACGTCGCGGACCGCGCGGCGCCTCGCGACACTGCTGGCACCGTCC   | 807711                            |
| QY         | 11448 | TCGGCCTCCCATGACGTACGCCCTACTTGGACAGGCACTCGGACGGCGAGCAGTTCCACG                                       | 1207                              |
| Db         | 80772 | CGGCGGCGACACCGCTCACCGTCGTCGTTCCACACCGCGCGGCTCTCTCGACGACGCGTCC                                      | 808311                            |
| QY         | 1208  | CCGACGACCTACACCGCGCAATGCTCTGACTGGGCTGGGACTGTTACGCGCGCTCACT   | 1267                              |
| Db         | 80832 | TCGGCTCCCTACACCGCGAGCGCTTGGACACCGTCTCTCGGCGCCCAAGGCCACGCGCGCT                                      | 808911                            |
| QY         | 1268  | GGCCATCGATCGACGCGCGCCGAGACCAAGCGCTTCGCCATCGCGTGGATGACAACTGGA                                       | 1327                              |
| Db         | 80892 | GGCACTTCGACGAGGCGACCGCCACCTCGACCTGGACGCTTCGTCTCTTCGTCCG  | 809511                            |
| QY         | 1328  | AGTAGCGCGACGAGCTTCCCACCGACGATCGGACGGCTACAAAGGGGAGAACTCGA   | 1387                              |
| Db         | 80952 | TCGCGCGCACCTTCGGAGCGCCGAGACAGGCGCAACTAGCGCGCGGCAACGCTTCCTGG  | 810111                            |
| QY         | 1388  | TCGTCGCGAGCTCGGCTCGC--CCGACAGCTTGGCGGCTGGTACACCTCTCTGAGCAC   | 1445                              |
| Db         | 81012 | ACGCCCTTCGCGCGCGCGCGCGCCACCGGCTTCGCCGCACTTCCCTCGGCTGGGGCC  | 810711                            |
| QY         | 1446  | CCCCGTGGCAGCGTGCAGAACTAGTTCACCGCCACCACTACCTTCCCGGACCGGACCGT  | 1505                              |
| Db         | 81072 | CGTGGACCCAGAGCTCGGCGATGAC--AAGCAGCGTGTCCGACCTCGAGCTCGAGCGCAT                                       | 811291                            |
| QY         | 1506  | CGACGCGAGCGCGTCTCGCCATGGAACGGAGCGGCATACGAGATCGAGCTCGCATCGC   | 1565                              |
| Db         | 81130 | CGCCCGTTCGCGCATGCGCCGCTGACCTTGGAAACAGGGCAACCGGCTTCTTCGACGCGGC                                      | 811891                            |
| QY         | 1566  | CTGGGACACCGCAGCAAGCTCGGCATCTCGTGGCGCGCTCC--CCGACGGAACCG  | 1622                              |
| Db         | 81190 | CTGCGCGCGGCGCGCGCGCTTCGCGCGCTCGACCTCGACCTGCGCGTCTCTGCGCAC  | 812491                            |
| QY         | 1623  | GCACAGCAACTCGGCAAGTAGCGAGCAGACCTGTACTGTGACGAGGACCCCTTCGACCT  | 1682                              |
| Db         | 81250 | CCAGGGCGACATCGCCCGCTGCTCGCGGCGCTGATCGCGCACCCCGCTGCGGCGCACCGC                                       | 813091                            |
| QY         | 1683  | CGCGGGTACTCTCGCTCGCCCTACTCGGAGCG-----CGCGCCCATCGACCCCGG  | 1736                              |
| Db         | 81310 | CGCCAGGTCTTCGAGACCGCGCGAGCGGCTCGCCAGCGGCTCGCGCGGCTCGAGCGCGC  | 813691                            |
| QY         | 1737  | CGCCCGATCCGTGCACCTCGCATCTCTCGTGGACACCCAGAGCGCTCGAGGCTTCGTCAA                                       | 1796                              |
| Db         | 81370 | CGCCCGCGGGAAGCCCTCTTGGAACTCGTCGACACCCAGATCGCCAGGCTCTCGGCA  | 814291                            |
| QY         | 1797  | CGCGCGGCACACGTCGTCTCCAGAGGTCCACTTCCGCGAGGCGGACGCGGAATCTC   | 1856                              |
| Db         | 81430 | CGCGAGCGCCACCGAGGTGGAGACGGCGCGCAGTTCGAGGACCTTCGACTCCCT   | 814891                            |
| QY         | 1857  | GCTCTACACGAGCGGCGCGCGCACACTTCACGGGCATCTGTCGTCGG  | 1904                              |
| Db         | 81490 | CACCGCGTCGAATCCGACAGCCCTGAACACCGCCACCGGCTCG  | 81537                             |
| RESULT 9   |       |  |                                   |
| AX211739   | LOCUS | AX211739   | 125401 bp DNA linear PAT 06-SEP-2 |
| DEFINITION |       | Sequence 35 from Patent W00159126.   |                                   |
| ACCESSION  |       | AX211739   |                                   |
| VERSION    |       | AX211739.1   | GI:15523950                       |
| KEYWORDS   |       |  |                                   |
| SOURCE     |       | Streptomyces noursei.  |                                   |
| ORGANISM   |       | Streptomyces noursei.  |                                   |
|            |       | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  |                                   |
|            |       | Actinomycetales; Streptomycineae; Streptomyces   |                                   |
| REFERENCE  |       | 1 (bases 1 to 125401)  |                                   |

|             |   |       |  |  |
|-------------|---|-------|--|--|
| AUTHORS     | Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T., Strom, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and Gulliksen, O.M.   |       |  |  |
|             | Gene cluster encoding a nystatin polyketide synthase and its manipulation and utility   |       |  |  |
| JOURNAL     | Patent: WO 0159126-A 35 16-AUG-2001;  |       |  |  |
|             | Norges Teknisk Naturvitenskapelige Universitet (NO); STIFTELSEN IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO); ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich (NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO); Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Valla, Svein (NO) |       |  |  |
| FEATURES    | Location/Qualifiers   |       |  |  |
|             | 1..125401   |       |  |  |
| source      | /organism="Streptomyces noursei"  |       |  |  |
|             | /db_xref="taxon:1971"   |       |  |  |
| BASE COUNT  | /note="ATCC 11455"  |       |  |  |
|             | 15664 a 49692 c 42871 g 17174 t   |       |  |  |
| ORIGIN      |   |       |  |  |
|             |   |       |  |  |
| Query Match | 7.4%; Score 147.6; DB 6; Length 125401;   |       |  |  |
|             | Best Local Similarity 45.4%; Pred. No. 4.6e-08;   |       |  |  |
| Matches     | 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;  |       |  |  |
|             |   |       |  |  |
| QY 8        | ACCCCACTCCCTCGACGACACCGTCCGCTACCGGCGACCGCCGCGCCGACTGC   | 67    |  |  |
| DB 80064    | AGGGCGGACACCGACCGCGCGCTTCGGACTGACACCGGCCCTGCTGGACCGCGCAC  | 80123 |  |  |
| QY 68       | TCCTCAGCTAGACGGGCGCCCTCTCGAGGTCTTCGTCGGGACGCTGAGGGACTGCGT   | 127   |  |  |
| DB 80124    | AACACGCGCGGCTACGCGGACCTCGGCGCCATACGCGCGCGCTTCCGCTTCCGCT   | 80183 |  |  |
| QY 128      | CGAACCTGTCTTCCTGGGGCGCGGCTGTACCGCGAGCCCTCGAGCGACGCGCCAG   | 187   |  |  |
| DB 80184    | GGGAAGCGCTCTCGCTCGCGCGCGCGCCACACCGTCCGCGCGCGATCGCCCGCG  | 80243 |  |  |
| QY 188      | GAACCTGTACGTGACCGCGATGAGTGTGAGCGCGCCAGCGATGTGACGCGCCCTGAAC  | 247   |  |  |
| DB 80244    | CGGCGAGGACACCGTACACATCGCGCTGTACGACCGCGCGCGCACCGTCTGTCG  | 80303 |  |  |
| QY 248      | CTGCGCGCTTCTGGGCTGACGAGCGCTCCACCGCGAGCTCTCTTCTACCGTGC   | 307   |  |  |
| DB 80304    | TCGACTCCTGTCTCCCGGAGGTCCCGCGACGACCGCGCGCGCGCGACCGTCC  | 80363 |  |  |
| QY 308      | CGAACAGGGTGGACGCTTCTGTGCGCCACCGCTCCACGAGAGAACAGCAATGACGC  | 367   |  |  |
| DB 80364    | ACCGGACTCCCTCTTCCACGTGAGTGACCGCGCTCCAGGGCGCGCGCGCGCAC   | 80423 |  |  |
| QY 368      | CGGCCATCTACCGCGCGCTGCTCCAGGAGCGCGCGCGCGCGCGCTGATCT  | 427   |  |  |
| DB 80424    | CGGCCA-----CCGTGCGGTCTCTCGGCCCGACCGCGCGCGCGCACCC  | 80474 |  |  |
| QY 428      | TCGGCGGTGTGTGCGCGCTCGAGCGCGCGGATCGCGCTCGGCTCGGTGCGGTCT  | 487   |  |  |
| DB 80475    | TCGCG-----CCACCGCATCCGACACCGCGCGCGCGCGCGCGCGCTCG  | 80525 |  |  |
| QY 488      | ACCATATGACGCGCGCGCGCTGCTGTGACCGCGCGCGCGCGCGCGCGCGCG   | 547   |  |  |
| DB 80526    | CGACCGCGAAGGCGCGCTGCGCGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTG  | 80582 |  |  |
| QY 548      | GCGCTTACAGTGTACTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG   | 607   |  |  |
| DB 80583    | GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG   | 80642 |  |  |
| QY 608      | ACCGGACGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG   | 667   |  |  |
| DB 80643    | AGTGGCTCGCGGACGCGCTGCGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG   | 80702 |  |  |
| QY 668      | CGGACTTCCCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG  | 727   |  |  |
| DB 80703    | CGGAGGACACCG  | 80762 |  |  |
| QY 728      | GCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG   | 787   |  |  |
| DB 80763    | AGAACC CGGCGCTTTCGCGCTCTCTGACCTCGCCCGCGACACCGCGCGCGCGCGCG   | 80822 |  |  |
| QY 788      | AGCAGTACCTTACTGTGAGACGAGCGGGTTCAGTTTACCGCTCCCGCGACCGCG  | 847   |  |  |
| DB 80823    | AGACCTGGCCACCGCCCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG  | 80879 |  |  |
| QY 848      | TCATCTCAACACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT   | 907   |  |  |
| DB 80880    | GCACGACGTGACG   | 80939 |  |  |
| QY 908      | GTTTCGCGACCCCAAGATCTGAGTGGGACACCGCGCGCGCGCGCGCGCGCGCGCG   | 967   |  |  |
| DB 80940    | GGAAACCGGACG  | 80999 |  |  |
| QY 968      | GAGGACTGGGTACGCGCGCGTTCACACCTCGCGCGCGCGCGCGCGCGCGCGCGCG   | 1027  |  |  |
| DB 81000    | CCGCGCACCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG   | 81059 |  |  |
| QY 1028     | GCAACTTGACTACCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG  | 1087  |  |  |
| DB 81060    | CGCGCGCGGACG  | 81119 |  |  |
| QY 1088     | TCACCGCAGACG  | 1147  |  |  |
| DB 81120    | ACATCGCGCGCTGCG   | 81179 |  |  |
| QY 1148     | TCGGCTTCCCATGAGTACGCGCTTACTGGACAGGACACCTGGGACGCGCGCGCGCG  | 1207  |  |  |
| DB 81180    | CGCGCGCGGACCGCGTACCGCTGCTGTCCACACCGCGCGCGCGCGCGCGCGCG   | 81239 |  |  |
| QY 1208     | CGGACGACTCACCG  | 1267  |  |  |
| DB 81240    | TCGGCTTCCCTACCG   | 81299 |  |  |
| QY 1268     | GGCCATCGATCGACG   | 1327  |  |  |
| DB 81300    | GGCACCTGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG  | 81359 |  |  |
| QY 1328     | AGTACCG   | 1387  |  |  |
| DB 81360    | TCG   | 81419 |  |  |
| QY 1388     | TGTCGCGGAGTGGCGGCTGCG--CGGACGCGTGGCGGCTGGTACACCGCTCTGAGCAC  | 1445  |  |  |
| DB 81420    | AGCGCTTCG   | 81479 |  |  |
| QY 1446     | CCCGTGGGCGGCTACGAACTACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG  | 1505  |  |  |
| DB 81480    | CGTGGACCGGCGGCTGCGCATGAC--AAGACGCTGTCCGCGCTGCGCGCGCGCG  | 81537 |  |  |
| QY 1506     | CGACGCGCGCGCGCTGCGCATGGAACGAGCGCGCATACGAGATCGAGTCCGACATCGC  | 1565  |  |  |
| DB 81538    | CGCGCGCTCGGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG  | 81597 |  |  |
| QY 1566     | CTGGGACACCGCGAGAGTGGGATCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCG  | 1622  |  |  |
| DB 81598    | CCTGGCG   | 81657 |  |  |
| QY 1623     | GCACGGAACATCGCAAGTACGAGCAGACCTGTACGTGCGCGCGCGCGCGCGCGCG   | 1682  |  |  |
| DB 81658    | CGAGGCGGATCTGCG   | 81717 |  |  |
| QY 1683     | CGCGCGGATCTGCG  | 1736  |  |  |
| DB 81718    | CGCGCGGATCTGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG  | 81777 |  |  |
| QY 1737     | CGCGCGGATCTGCG  | 1796  |  |  |
| DB 81778    | CGCGCGCGCGGAGCG   | 81837 |  |  |
| QY 1797     | CG  | 1856  |  |  |







Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
1 (bases 1 to 12744)

REFERENCE  
AUTHORS  
TITLE  
Schauwecker, F., Pfennig, F., Grammel, N. and Keller, U.  
Construction and in vitro analysis of a new bi-modular polypeptide  
synthetase for synthesis of N-methylated acyl peptides  
Chem. Biol. 7 (4), 287-297 (2000)

JOURNAL  
MEDLINE  
20243879  
10780924

REFERENCE  
AUTHORS  
TITLE  
Schauwecker, F., Pfennig, F., Grammel, N. and Keller, U.  
Direct Submission

JOURNAL  
Submitted (12-NOV-1999) Institut fuer Biochemie, Technical  
University Berlin, Franklinstr. 29, Berlin 10587, Germany

FEATURES  
source

Location/Qualifiers  
1. .12744  
/organism="Streptomyces chrysomallus"  
/strain="ATCC 11523"  
/db\_xref="ATCC:11523"  
/db\_xref="taxon:1899"  
1. .12744  
/gene="acmC"  
1. .12744  
/gene="acmC"  
/note="ACMS III; peptide synthetase"  
/codon\_start=1  
/transl\_table=11  
/product="actinomycin synthetase III"  
/protein\_id="AAP42473.1"  
/db\_xref="GI:7228478"

/translation="MTSSGLEDLPLSPMOEGLLFHSRYEQDGDADVAVQHVLDLAGP  
LDGARVAGLVORHPNRAAFROVDSTVOLTSRHVDLPWEEDLSMFAPEAE  
AELSRIAAEKRGFRDPAEPFLRTVTAEEHRLILTHHLLDGGSTPILVREL  
TALYADGPTEAPRVPTPYROYLGWLAQODRPAEAAWREALGDLQPTLVAPDPAAR  
ALMPERTTELAEARTAAALVEMARRHGVTLNTVLOAAWGLVLSRRTHGHVDVGAAR  
GRDQLAGVESVGLLITWVPVVRDLDAQSLLEAVRRLQDEQSLTSHHGLIARIQ  
QLTGMSELDFTSLFENIPWQDPAELPATGLRITPDLGRDRATHPTLLTIAAPGRKL  
YLRDLVDYDRAEACALVERILVLDLAVADAERPLGRVLDLITAEHRELLHAGNG  
AGMLPEASLPFLFRAVAAADAVAVEGEALRYRELDERANRLAHLAAHGVGE  
RIVALALPRSVLDVAVLAVKAGAAVLPDPEYPANRLAHMVTDAQPTLVASTETE  
AKLPDRHGLDDPAVADLADSGRPAHPVLPDHPAIVITSGSTGPKGVMPA  
GGLNLQHHRRVAGDEPGTRTAQTAISFDVSAQEVLSVAFGLTIVIPDEVRRA  
ARFAGMLDRQVDELAPNLVLEALAEAVETGRTLPQLRTAQAGAGLTLISRTAF  
HRSAPGRLLHNGVPTTHVVTVAHALGDPEDEWRLPAPIGRPIDNTHAVYRTRAVRL  
EPGVGELYTAGLAGLARGYLRPAALTAERVADPYGLEPGRMYRTGDVLRVADDEL  
EFCGRADHGVKRGFRIEGEEIEKVLTDHPDIAQAAVVTPRHPGDTLRVAVYGRE  
LRPEQVREFTRRLPHMVPAAVQLERLPLTPNGKLDRAALPEPDALAGTGREART  
POQIYCDLPAQVLGLPWGVGDDDFELGCHSLIAFLIARIAAESVELGLTLEA  
RTAAVAALHDLGAPARTALTAKQLDPAVLSFAORLWFLHMKEGSPATYNIPLAIR  
LTGELNLAALEADYTGHRSLRTVFFVEDGTPPYORLDTATIELVRTATTEAELT  
DHLLRAARHPFDLAGEPPLRTELFELSAAREHLLVVVHHITAGDWSLGLSLARAY  
AARAEGRAPQWAPLPVOYADYTLQWELLGDNDPDSLFATQVAYWTTETLAGLDQIT  
LPDRPRPAVMTYRGDYLTVDIDPELHRLRTELARGSGASLFWLQAGLAALIKRLGA  
GDDIPLGSPITAGTDOALDILIGFEVNTLVLRDITDPTFTOLLTRVRETSLLAAAH  
ODVPFVLEVLNPTRTLAHPLFOIMALQNAPEGTFPLGLTVDVAPORTGAKED  
LFTSLAERRGAGEPGOGITGAVESSDIYDAPITQALFNWHLHLLDTATQAQPOPLSR  
IDLTLAQEHQLDITWDTAVEVGPDLPLPAFRAQAATPOAVALLTAGDSDLSYAEID  
ARANHALHALLREGAGDRVALALPRTAEVLVALLKTAAYLPDHPDPAARLS  
HVLGDARPAALLDITTEQHLPADADTRLALDSAEVALLADCPDTPAEAGSVTPAP  
GSAAYVITSGSTGRPKGVVPHSALVNFVTAMRQAPLPQERLLAVTAFDIAL  
ELYHPLLSGNAVLAKEAVPQPSAVLDLIRAGVTVTQGTPSLWQLLVGHDAELRG  
LRMLVGEALPLSLAELRALTDLLNLYGTETTWSTAAELAGGTGAAPIGRPTAN  
TRYVYLDLQGPVAGVWGLYTAGLAGLARGYLRPAALTAERVADPYGLEPGRMYR  
TGDLWRNPDPGELEFVGRADHGVKRGFRIEGEEIEKVLTDHPDIAQAAVVRDPOG  
DARLAVYTVGSGSADARDEQVODQLEMODLYDSVYTAASETAFGENSWNSVDG  
LPDLOQMRERDPTTVIRGLNPRDFVLEIGVGTGLLSRLAPHCEYNGWGTDSPTVI  
ADURGHVEADPELAARVQLRTQPAHFDQLPHGFDTVLVNSVVOYFPNAGYLEQVID  
HALUKIAPGTVFIGDIRNPRLLRTTAVHTARATDPAHAAVRHAHEHALYKELER  
LIDPFTALGHVRPLAGVDIRLKPGTVENELTRYDIRATLHKAGATHPLGCPPER  
PWSGHADPDLAALAROLDSEPARLVTVGINPRALRELAHQALDATATAPAS  
TADAGQPDPLDALHGRHGYWTAATWNTHRPDADVLVVALSDLDGATPTGTVAS  
AGTSTPATPLSLTTPNAGRGTSALLTLTREHARTHLPDQMPSALVPLDLRLPTAN  
GKLDRAALPAPDFTLAGTGREPTPOEQIVCDLFTOVGLPRVGIIDDDFELGSHSL  
ATRLIAQIRALFVLELRSLFEGPTPAVAAARLDTAGPGRALTVOERPATMPLSFA

gene  
CDS

BASE COUNT 1846 a 5446 c 3884 g 1568 t  
ORIGIN

Query Match 6.08; Score 119.8; DB 1; Length 12744;  
Best Local Similarity 44.68; Pred. No. 9.6e-05;  
Matches 558; Conservative 0; Mismatches 687; Indels 6; Gaps 2;

Qy 583 AACACGCGCCCGCGCGTGGGACACCGACGACGACGCGCGTCCCTTACACGCAC 642  
Db 9049 AGCGACGCGCCACGCTCAGCTACTCCGAGCTCAACACGCGCCACCTCCGCCAC 9108  
Qy 643 CACGCGACCGTATGCCCGCTCGGCCCGACTTCCCGGTGTGTCGGGTGCGCGTCTGTC 702  
Db 9109 CAGCTCACCGCGCGGATCCGCCCGCGACGCGCGTCCCTCTCTCAACGCTCC 9168  
Qy 703 GGCACCGGAACACGCGAGGTTCGCGCGCGCGCGTCTCGCGCTCGCGACCGCGC 762  
Db 9169 CCGACACCGTACACCGTCTCGCCCTCGCGACGCGCGCGACCTACATCCCCCTC 9228  
Qy 763 ACCGACGCGTCCGCAAGTACGAGGAGTACCTCTTACTTGTGCGACGCGCGGGTTC 822  
Db 9229 GACAGCGCTACCGCGCGCGCTACCGCTCGCTCTCGACGAGACCGCGACCAACTC 9288  
Qy 823 ACGTTCACCGCCC ---TCGCCGACCGCTCATCTGCTCAACACCGAGGTTCGCCGCCAC 879  
Db 9289 CTCATCACCGACACACCGACCTCGACACCAACCAACAGTTCACACCGCGCGAC 9348  
Qy 880 ACGCGCGCGAGATCGAGAACCGCGAGTGGTTCGCGACCGACCGCAAGATCCACTGGGACAC 939  
Db 9349 ACCCGCGAGCGGGAAGACCCCGGGAACCGGGAACCAACACCGACCGCGCGACCGC 9408  
Qy 940 CCGCGCGGAGATGGTGTGCGTGCATCGGACGACTTGGCGGTACGCGCGGTCTTACACCTCG 999  
Db 9409 GCCTACATCATGTACACGCGGCTCCACGCGCGCCCAAGGGCGGTCTATCGCCACCCAC 9468  
Qy 1000 CCGAACCTCGCGGACTGAGACACTTGGCGGCAACTTTCGACTACCGGAACACCGCTCGCG 1059  
Db 9469 CGCAACATCACCGGCTCGCGCTCGACCGCGCGCTTCGACCGCGCGCGCGCGCTC 9528  
Qy 1060 GGCATCGAGTCCCGCGCTCTCGAGATCAACCGGACGAGCGGACGCGGACGCGCTGGGTG 1119  
Db 9529 CTCCTCCACTCCCCACCGCGCTTCGACCGCTCCACCTACGAGATCTGGGTCCCCCTCCTC 9588

|   |        |
|---|--------|
| cedex,  | FRANCE |
| Location/Qualifiers   |        |
| 1. .22449   |        |
| /organism="Streptomyces pristinaespiralis"  |        |
| /strain="SP92"  |        |
| /db_xref="taxon:38300"  |        |
| /clone_lib="PIBV3 cosmid"   |        |
| 115. .7890  | gene   |
| /gene="snbc"  |        |
| 115. .7890  | CDS    |
| /gene="snbc"  |        |
| /codon_start=1  |        |
| /transl_table=11  |        |
| /product="Pristinamycin I synthase 2"   |        |
| /protein_id="CAA72311.1"  |        |
| /db_xref="GI:2052248"   |        |
| /db_xref="SPTREMBL:O54959"  |        |
| /translation="MLVDDARLPLTAAGLGIFQAQDPHSPVNTGEYVELTGGLD<br>ADRFEAALRATVARTDALRLRFTDAEGTQVSADVDFTLHRVDVSAEDDPFAAAAD<br>AMRDIARPLDVTCTGRLTQALFTLPGRHRWFQHVHLLLDGDFGLRVAERAAA<br>YRAGAEAGDAEAGDEPGIRDLAAEADMRASDAYAQRHAWLATPHPTVTLGATTA<br>PCAGALRRTASAGTEVLAGLDAAVARAEQVAFPAAVAAVOWLTGERTVTLGMP<br>VTARTSAARRAPGTVANYPALHVRPAPLAELVASVRDALEPLLRHORYPYEAR<br>RELGAVERAYCPVNVMPYGRALDFGPAASVHPYSTGVEDLKFSEYEEBGGGAR<br>IDLEANAAYIGSETDAHLAREVRELGAFAEAGPTPIGRIHAAPAAAPAAAPITGG<br>PALHTLFAARAAATPHAVASHDGHLYAQDLARNLARHLRVGOGAPERLVALTL<br>PRSLDLVALLAVLKSGAAYLPDIPDYPADRLAWLMEADAPVLVDPOAMERTLAGYP<br>DTAPRTAGWEIHAIVYITSGSTGRKGVVIPSNNVRLFTSTDHWFSGPDDVMTLF<br>HSYAFDSWEIHALHGGRLVVPVYHVSRSFGDFDLRLAREKVTNLNOTPTARHOL<br>DAADRFTAAPAEALURYVVFGEALDVARLADYARRGTAARLNVNYGITETTVHVT<br>ALTPGPHAVPCTPSLLGGPIDLTPRVLDDALRLVPVPGFTGELYVAGLAGLARGYLNR<br>PLTAAHPGTAQAQAAVVRREDRHGLRLAAVATPCTGEPDVAGNRALLAGOLPAHSLPAS<br>VLTHLALPLVANGKLDRAALPAQAPAPETADAPAPADPGEARLRRIFADFLGTGPRV<br>GADSDFFTLGGHSLLAGRLIAAVRDTLCAEITIQDLFEAPTAPGLAARVTAATGRPP<br>VRDRHGDGPTSYAQRHMVMVLRVEGAAPTYNMLPALARLUSGTLNRLTAARLYXDVY<br>LVHESLQTPVYPRGEHGVTPRLLPPEQAPRWIESRVDDDEOLPAALTAARAREPFLNDCO<br>LYIRAHFTLPLGGDSVLLMLHHVAGDGSRLPLGLDLSITAYARLAARAPDPPLPY<br>SYADFAVWORELLGDHPDLSLAAROLAHWESTLAGLPDQLELPADFPRPVVAERHGG<br>SVPLAVPQLHARLDTFAAHGATCMHVLQAAALALLTRIGAGTIDIPICGTASRPDS<br>ALDLGVFFVNPVLRADTSGDPTETELLARVRATALAAYAHQDVPFVRVERLOPER<br>SAAHRLPFVVLVYQDFEARHLKGLDRREDRDFGAFDILTFNLVHRHGGDEEFAG<br>IDGLEYDTALYERATAELTGRLLRLDLDAAVVTDPQAPVSGHLLLEEAHARLAAPFN<br>TARVPYRAGLAELFTQARRTPDAVAVVGDGTELTAYELDARSALARRLAAGACGR<br>ESAVAVLOORSVHLVSLVALVKGGFVPLNTRYPADRMRLITADKARVLULTDRAIT<br>DTAHECRSWAGDADVLVDELPTGCEPLVPAPHOLAYVMYTSGSSTGNPKGVATTH<br>GDVALAADTCHRTGNQORVLLHSPYSFDTSQYELWVFLSGGTVVAPPDGLDTRAL<br>ERAIVDGGVTGTLFSLGFLNLAAESPHCFCTGREVMTGGDVVSPAARVLUASPSPT<br>MYADGGPTETTTATTHFMRAPOQOETTVPIGTPLDNTTCHVLDLRDPVPVGTGTE<br>LYIGGSLARGLYNRPDLTARFVADTPAGGMYRTGCLVLRDRDGYFTLGRADHQC<br>VKVGRFTELGFIESVLGAHPCIAQASAVLVREDRPGDKRLVAYVVPAGCATADATVLR<br>RHVAALPDYMWPAFVPLDALPLTGNCKLDRRALPVPDFGNAADSAGRAPRTPRELL<br>AVLFAEVGIEKESIDDSFFDRGDSIMVLIQLVRRARRAGLLNLAALFHEKTVQAFLA<br>PVATLADTTREEPAAVATGDVPLTPIIMHMLRERGSPVDAFSQTMATVKTGLDRER<br>LVAALTLIDHLDALRAVFSRDDADWTMRIEAPGTVDAAWMVHRVHDAGLDSQALAAAL<br>TTAAEAARERLAPFAGGGOAVLWDGPHRGRULLTLHLHVVDGVSWRILLPLKLA<br>AWEAATPGSAAPALPPVTTSPRAVAHLEQQOAGCPERAEALATLGMLEAAGSPVTAQ<br>PLDGLRDTAANTARGTLTLDITTTTAAALLTAPAFNGRVDDILLTAPALAVTDWRKRR<br>GLGDSDAVLTDVEGHGREPSTGSAELSTVGVFTSLYIPVRLAIGTLTDRADALAGPAA<br>GTAVKRIEALRQIPDGGIGHGLRHLDPRTCRSLAALAPYPTPLGPNVYLRGFRFDEP<br>ADGWSMAAPDAEAIIGDGTDPRLPLAHAVEVNAITLDEPAGPKLATWTFAGGLLPAE<br>VRDLGCTWFRALNALVTHAERPEAGGITPSDLISVLQSQDQIDQLEAAKWVsp" |        |
| 7887. .22433  | gene   |
| /gene="snbc"  |        |
| 7887. .22433  | CDS    |
| /gene="snbc"  |        |
| /codon_start=1  |        |
| /transl_table=11  |        |
| /product="Pristinamycin I synthase 3 and 4"   |        |
| /protein_id="CAA72312.1"  |        |
| /db_xref="GI:2052249"   |        |
| /db_xref="SPTREMBL:O07944"  |        |
| /translation="MTRSGRLADVLPSPLOQGLLTFQVYGAETHGCPDYTVTQVMVFEFLR   |        |



```
QY 861 CGAGGTCGCGCGCCGACACGCGCGCGCGGAGATCGAGACGCGCGAGTGGTTCGCGGACC- 919
Db 1671 CGCGGCTCTCAAAATCGCGCGCGCTTACCTACCTCGCGCTCGACCCGCGTACCGCGCGACCG 1730
QY 920 ---T-CCAAGTCCACTGGGACACCGCGCGGAGAAATGGTCTGCGTCATCGGACGACT 974
Db 1731 GCTGGCTGTGATGCTCGAGGACGCGCGCGCGCTTCTGCTCTCGACCCGAGGCGATGAC 1790
QY 975 GCGGTAGCGCGCTTCTACACCTCGCGGAACCTCGCGGACTGGACACCTTCGCGCGCACTT 1034
Db 1791 CGAGGACCTCGCGCGCTACCGGACACCGCGCGCTACCGCGCTGCGCGCGCCACCC 1850
QY 1035 CGACTACCGAAGACGCGCTCGCGGCGATCGAGTGGCGCGGACTGTTGAGATACCGC 1094
Db 1851 CGCGTAGTCTATACACCTCGCGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1910
QY 1095 AGACGGGACACGCGCTGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1154
Db 1911 CACGACGCTGGTGGCTGTTCACCTCCACCGGACCACTGGTTCGGCTTCGCGCGCGCGCA 1970
QY 1155 CCCCATGAC---GTAGCGCTTACTGGACGACGCGCTGGGCGGAGCAGTTCACGCGCA 1211
Db 1971 CGGTGACCGCTCTCCACTCTTACGCTTTCGCTTTCGCTGGGAGATCTGGGCGCG 2030
QY 1212 CGACTACCGCGCGATGGCTCG------ACTGGGGCTGGGACTGGTACGCGCG 1259
Db 2031 CCGTCTGACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2090
QY 1260 CGTCACTGGCCATGATCGACGCGCGCGCGGAGACCAAGCGCTCGCCATCGCTGGATGA 1319
Db 2091 CTTCTCTGACCTGCTGGCGCGCGGAGAGGTCACCGCTTCAACGACGACACCGCGCGCT 2150
QY 1320 CAATGGAAGTACGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCTTACACGCGCA 1379
Db 2151 CCACGAGCTCGACGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCTGAGTA 2210
QY 1380 GAATCGATGCTCGCGGAGCTGGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1439
Db 2211 CGTCTGCTGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2270
QY 1440 GAGCACCGCGTGGCAGCGCTGACGACTACGTCACCGCGCGCGCGCGCGCGCGCGCGCG 1499
Db 2271 CGGACGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2330
QY 1500 GACCGTCTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559
Db 2331 GCACGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2390
QY 1560 CATCGCTGGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1616
Db 2391 CATCCCGGACCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2450
QY 1617 AACCGCGCACACGATCGGCAAGTACGAGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1676
Db 2451 CACCGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2510
QY 1677 CGACCTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1736
Db 2511 GCTGACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2570
QY 1737 CGCGCGGATCGGTCACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1796
Db 2571 CACCGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2630
QY 1797 CGCGCGGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1856
Db 2631 CGACGAGTCAAGTCCGCGCGCTTCCGATCGACTCGCGGAGATCGAGACGCTCTCGC 2690
QY 1857 GCTCTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1916
Db 2691 CACCGACCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2750
QY 1917 GCGCAT 1922
```

Db 2751 GCGGCT 2756

RESULT 14  
SPSNBGEN  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

SPSNBGEN 22449 bp DNA linear BCT 24-MAR-1997  
S.pristinaespiralis snbc and snbde genes.  
X98690.1 GI:1906377  
Pristinamycin I synthase 2; Pristinamycin I synthase 3 and 4; snbc  
gene; snbde gene.  
Streptomyces pristinaespiralis.  
Streptomyces pristinaespiralis  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 22449)  
de Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S.,  
Famechon, A., Bamas-Jacques, N., Crouzet, J. and Thibaut, D.  
Pristinamycin I biosynthesis in Streptomyces pristinaespiralis:  
molecular characterization of the first two structural peptide  
synthetase genes  
J. Bacteriol. 179 (3), 705-713 (1997)  
97158664  
9006024  
2 (bases 1 to 22449)  
de Crecy-Lagard, V.  
Direct Submission  
Submitted (21-JUN-1996) V. de Crecy-Lagard, Institut Pasteur, 25  
Rue de Dr. Roux, Paris Cedex 75724, FRANCE  
Revised by [3]  
3 (bases 1 to 22449)  
de Crecy-Lagard, V.  
Direct Submission  
Submitted (24-MAR-1997) V. de Crecy-Lagard, Institut Pasteur, 25  
Rue de Dr. Roux, Paris Cedex 75724, FRANCE  
On Mar 25, 1997 this sequence version replaced gi:1483602.  
Location/Qualifiers  
1. .22449  
/organism="Streptomyces pristinaespiralis"  
/strain="SP92"  
/db\_xref="taxon:38300"  
/clone="piBV3 cosmid"  
115..7890  
/gene="snbc"  
115..7890  
/gene="snbc"  
/codon\_start=1  
/transl\_table=11  
/product="Pristinamycin I synthase 2"  
/protein\_id="CAA67248.1"  
/db\_xref="GI:1483603"  
/db\_xref="SPTREMBL:Q54959"  
/translation="MLVDDARLPLTAOLGIWFAQADPHSPVFNFTGVVLTGCLD  
ADRFEARLRATVARTDALRLRHTDAEGTVTSVADDDVFTLHRVDVSDDEPAAAD  
AMRDVIARLDVTCGTRTQALFTLGGRRHWFQHVHLLDLDGFGFLRAVERVAAA  
VRAGAGDAEAGDEPGIRDLAAEADMRSADYADQRAHWLATPHPTPTVLAGTTA  
PGAGALRTASAGPEVLGADAVARAETGRQVAFPAAPAAVYVOMLTGERTVTLGMP  
VARTSAARAPGTVANVPALHVPAPAAPLAELVASVRDALELLRHORYPYEAR  
RELGAVERAYGPVVNMPYGRALDGGPAAASHPVSTGCPVEDLKFSVVEEDGSGAR  
IDLEANAAYGESETDAHLARFRELGAFAEAGPOTPLGRHIAAAPAAAPAAPTTGG  
PALHTLFAARAAATPHAVASHDGHLYAQDLARANLRAHLRLEVOGAPERLVALTL  
PRSLDLVALLALVLSGAAYPLDPDYPADRLAWLEDAAPVLDLPQAMTBDLAGYF  
DTAPRTAVDGAHPAYVITSGTGRPKGVVIPSNNVRLFTSTDHWFGRGPDVMTLF  
HSAFDSWEIWLGAHLGGRLVVVPHVSRSPGDFLDLLAREKVTVLNQTAFPHOL  
DAADRATAAPELALRYVFGGELDARLADMYARRGTAARLVNMYGITEVTVHVT  
APLFGCHAVPTPSLLGGIPDLTTPRLDAALRPVPPGFTGELYVAGAGLAGYLNRP  
ALTAQRFPADPYGAPTRMPTGDLVRLDDGTIAYLGRGDQVKIRGRIEELGEN  
ALTAQRFPADPYGAPTRMPTGDLVRLDDGTIAYLGRGDQVKIRGRIEELGEN  
VLATHPGVAQAAVREDRHGLRLAAYAPETTPGEPDVAGWRLAAGOLPAHSLPAS  
FTLLAALPLTANGKLDRAALPAPAPAPADEPADPADGEAALLRIFADTIGTPRV  
GADSDFTLGGHSLLAGLIAAVRDTLGAEITIQDLEAPTPAGLAARVATLRTGTRPP  
LVPRDHGDPGPTSYAQRHMWVLRVEGAAPTYNMPALRLSCTLNRTLRALYADV  
VRHESLRVYPRGEGVTPRLPLPPEQARFWVIESRVDDGLPALTAANAREPNLDCQ

gene  
CDS





|    |      |  |      |
|----|------|--|------|
| Qy | 333  | CCCACCCGGTCTCAGGAGAGAACACCAAGCAATAGACCGCGGCATCTCACCGCCGCGCGTGG-- | 390  |
| Db | 1131 | GGTGGGGAACGGGCGCTACGGCCCGCTGCTCAACGTCATAGCCCTACAGGCGCGCCCTCGGA   | 1190 |
| Qy | 391  | --CTCCAGGAGACCGGCGCGGAGACATCGCCCTGATCTTTCGGGCGGTGCTGTGCGCGCCT    | 447  |
| Db | 1191 | CTTCGGGCGGCCGGCGCGTCCGTCACCCGCGTGTCCACCGGCCCGCGTGGAGGACCTGAA     | 1250 |
| Qy | 448  | GCAGCCGGGCATCCGCTCCGGGCTCGCTCCGTCGCGCTTACCAACATGACGCCGCCCCAGC    | 507  |
| Db | 1251 | GTTCCTCGTGTAGAGAGAGCCGACGGCTCCCGCGCGCGGATCGACCTCGAGGCCAACGC      | 1310 |
| Qy | 508  | GGTGGGT-----CTGGACCCCCAACGCCCGGTACCAACCCAGCGCGCTACCAAGCT         | 560  |
| Db | 1311 | GGCGGTGTAGCGGAGAGAGACGACGCCCATCTGGCGGCTTCGTGCTCGCTTCTCTGGG       | 1370 |
| Qy | 561  | GTACTACCTGCATCCGACCAAGACACAGGCGCCCGCGGCTGGGACCAACGGGAGCAGCAC     | 620  |
| Db | 1371 | CGCCTTCGCGAGGCGGGCCGGACACAGCGCTCGCGCCGGATCCACGCGCGCGGCCCGC       | 1430 |
| Qy | 621  | CGACGGCGTCCGCTTACCGACACACGGCACCGGTGATCGCTCGGCGCCCGACTTCCCGCT     | 680  |
| Db | 1431 | CGCCCCGCGCGCGCGCGCCACACACCGCGCGCGCCGCTGTGACACCTGTTCGCGCGC        | 1490 |
| Qy | 681  | GTGTGTCCGGTCTGGCGGTCTGGGACCGCGACCGACGACGAGGTTCGCGCGCGCGCGGT      | 740  |
| Db | 1491 | CCGCGCGCGGCCACCCCGCACCGGTGSCGCTCAGCCACGACGAGACACCTCACCTA         | 1550 |
| Qy | 741  | CGTGGCGCTCGACCCAGCCAGCAGCGGCTCCGGAAGTACAGAGACAGTACCTCTA          | 800  |
| Db | 1551 | CGCGCAGCTCGACGCGCGCGCAACCGCTCGCCACCGGCTCTCGTAGAGAGGCGCAGG        | 1610 |
| Qy | 801  | CTGTGTGACCGACGCGCGGGTTACGTTTACGCGCTCGCGCCCGCCGCTCATCGTCAACAC     | 860  |
| Db | 1611 | CCCCGAACGGCTGTCGCGCCTCACCTCGCGCGCTCCCTCGACCTGTGTCGCCCTGCT        | 1670 |
| Qy | 861  | CGACGGTGGCGCGCACACACCGCCGCGAGATCGAGAACCGCGAGTGTTCGCGCGACC-       | 919  |
| Db | 1671 | CGCGGTCTCAAACTCGGGCGCGCTACTGTGCGCTCGACCCGACATACCGCGCGACCG        | 1730 |
| Qy | 920  | -----CCAAGATCCATGGGACACCGCCCGGGAGATGGGTGTGCGTCTATCGGACGACT       | 974  |
| Db | 1731 | GTGTGGCTTGATGCTCGAGAGACGGCGCCCGCTCTCTGCTCGACCCGCGAGCGGATGAC      | 1790 |
| Qy | 975  | GGGCTACGCGCGCTTCTACACCTCGCCGAACCTGGCGGACTGGACACTTCGCGCGAACTT     | 1034 |
| Db | 1791 | CGAGGACCTTCGCGGTATCCCGGACACCGCCCGCGGTACCGCGTGAAGCGCGCCACCC       | 1850 |
| Qy | 1035 | CGACTACCGGAACACGCGCTTCGGCGGATCGAGTGCGCCCGACCTGTTCGAGATCACCGC     | 1094 |
| Db | 1851 | CGGCTACGTCATCTACACCTCGCGCTCCACGCGCGCGCCCAAGGGCGTGTGATCCGCA       | 1910 |
| Qy | 1095 | AGACGACGGGACACGCCACTGGGTGCTTCGCGCCGACGATGACGCCCTACGGCATTCGGCCT   | 1154 |
| Db | 1911 | CAGCAACAGTGTGCGGCTTTCACCTCCACCGACCACTGGTTCGGCTTCGCGCCCGCAGCA     | 1970 |
| Qy | 1155 | CCCCATGAC---GTACGCTACTTGACAGGACCTTGGAGCGCGAGCAGTTCACAGCCGA       | 1211 |
| Db | 1971 | CGTGTGGACCTCTTCCACTCTACGCTTTCGACTTCTCCCTCTGGGAGATCTGGGGCGC       | 2030 |
| Qy | 1212 | CGACCTACCCCGCAATGGCTCG-----ACTGGGGCTGGGACTCGGTACGCGGC            | 1259 |
| Db | 2031 | CTGTCTGACGGCGCGGCTCGTCTGCTGCCGTACCAAGCTCAGCGCTCGCGCCGGGA         | 2090 |
| Qy | 1260 | CGTCACTTGGCCATCGATCGAGCGCGCCGAGACCAAGGGCTTCGCCATCCCGTGGATGAA     | 1319 |
| Db | 2091 | CTTCTCTCGACCTGCTGGCCCGCGAGAAGGTACCGTCTCTAACCGAGACACCGACCGCCTT    | 2150 |
| Qy | 1320 | CAACTGGAAGTACGGCGACGGAGCTGCCACACGACGACATCCGACGGCTACACAGGCA       | 1379 |
| Db | 2151 | CCACGAGCTGACGGCGGACAGGGGCCCGTACCGCGCCCGCCGAACTGGCCCTCAGGTA       | 2210 |
| Qy | 1380 | GAACTCGATGTCGCGGAGCTGCGGCTCGCCCGACAGCGCTGGCGGCTGTGTACACCTCTCT    | 1439 |

|    |      |  |      |
|----|------|--|------|
| Db | 2211 | CGTCGTCTTCGGGGGAGGCTCTCGACGTGGCCGCTCTCGCCGACTCGTATCGCCCGCG     | 2270 |
| Qy | 1440 | GAGCACCCCGTGGCAGCGGTGAGAACTACGTACCGCCACACACACACACACACACACACG   | 1499 |
| Db | 2271 | CGGCAGCGCCCGCGGCTGTCAACATGTACCGGCATACCGAGAGACCCGTCGACGTCAAC    | 2330 |
| Qy | 1500 | GACCCTGCAGCGGACGGCGCGTCTTCGCATGGAAACGCGCATACGAGATCGAGCTCGA     | 1559 |
| Db | 2331 | GCAGGCCCCCTCGGCCCGCGCACGCGCTCCCGGCGACCCCTCCCTGCTCGCGGCGCC      | 2390 |
| Qy | 1560 | CATCGCCTGGGACACCGCGACGAAGCTCGGCATCTCGGTGGCGCG---TCCCCCGACGG    | 1616 |
| Db | 2391 | CATCCCGGACCTCAACCCCGCGCTCTCTGGACCGCCCTTGGGGCCGTCGCGCCCGCGCTT   | 2450 |
| Qy | 1617 | AACCCGCGACACGAACATCGGCAAGTACGGAGCAGACCTGTACGTTCGACCGCAGGACCCCT | 1676 |
| Db | 2451 | CACCGGGAGCTGTACGTCCGGGGGCGGCTCGGCCCGCGCTACCTGAAACCGGCGCCG      | 2510 |
| Qy | 1677 | CGACCTCGCCGGGTACTCGCTCGCCCTTACTTCGGGAGCGCGCCGCCCATTCGACCCCGG   | 1736 |
| Db | 2511 | GCTGACCGCCACCGGTTCCTCCGCGACCCCTACGGGGCGCCGCGCACCGCATGTATCG     | 2570 |
| Qy | 1737 | GGCCGATCGGTGCATCGGGATCTGTGTGAACACCAGAGCGTTCGAGGTCTTCGTCAA      | 1796 |
| Db | 2571 | CACCGCGGACCTTGTTCGCGCACTCGACACGGCACGTACGCCTACTTGGGCCCGCGCGA    | 2630 |
| Qy | 1797 | CGCGGGCCACACGTCTCTCCACAGGTCCACTTCGCCGAGGGCGACACGGGAATCTC       | 1856 |
| Db | 2631 | CGACAGGTCAAGATCCGGGGCTTCGGATCGAACTCGCGGAGATCGAGAACGTCTTCGC     | 2690 |
| Qy | 1857 | GCTCTACACCGAGGGCGCCGCGACACTTACCGGCATCTGCTCGCGGAGATTGGCCA       | 1916 |
| Db | 2691 | CACCCACCCGGGCTCGCCACAGCGCCGCGCTGCTCGCGAGGACCGGCACCGCGACCT      | 2750 |
| Qy | 1917 | GGCGAT   | 1922 |
| Db | 2751 | CGCGCT   | 2756 |

|            |           |
|------------|-----------|
| RESULT 15  | JOURNAL   |
| AF210249   | MEDLINE   |
| LOCUS      | PUBMED    |
| DEFINITION | REFERENCE |
| ACCESSION  | AUTHORS   |
| VERSION    | TITLE     |
| KEYWORDS   |           |
| SOURCE     | JOURNAL   |
| ORGANISM   | MEDLINE   |
|            | PUBMED    |
|            | REFERENCE |
|            | AUTHORS   |
|            | TITLE     |

AF210249 77457 bp DNA linear BCT 29-AUG-2000  
Streptomyces verticillus bleomycin biosynthetic gene cluster,  
complete cds.  
AF210249 AF149091  
AF210249.1 GI:9937210  
Streptomyces verticillus.  
Streptomyces verticillus  
Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;  
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
1 (bases 74421 to 74693)  
Du, L. and Shen, B.  
Identification and characterization of a type II peptidyl carrier  
protein from the bleomycin producer Streptomyces verticillus ATCC  
15003  
Chem. Biol. 6 (8), 507-517 (1999)



```
polyketide synthase
Chem. Biol. 7 (8), 623-642 (2000)
20500448
PUBMED
11048953
4 (bases 1 to 77457)
Du.L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.
Direct Submission
Submitted (01-DEC-1999) Chemistry Department, University of
California at Davis, One Shields Avenue, Davis, CA 95616, USA
On Aug 29, 2000 this sequence version replaced gi:5326869.
Location/Qualifiers
1..77457
organism="Streptomyces verticillus"
/strain="ATCC15003"
/db_xref="ATCC:15003"
/db_xref="taxon:29309"
complement(1199..2185)
/codon_start=1
/transl_table=11
/product="Syrp-like protein"
/protein_id="AAG02342.1"
/db_xref="GI:9937211"
/translation="MTGSGVTLTPGGIIPRGRGELTTGAEDVGLGDLGADGDMVRAH
GRLRERLATDGLILLHCLPTDGGVDGDFHVDVSGDGPDPYTERSPRSVVKNIY
TSTETPADQIPMNENSYAHWFSTLYFFCHTADPTGGATPIADGRAVLDLIPAEVR
RFSQGVYTRFDMGLSWOEAFQTDGRDGRVHCHRAHGOEFSWDGDLVLRHHRP
ATAYDPTGAEFWENOHLFHPSSLDPLDQLVLTETGNGLPRDALFADGTPIPDAD
LATVRAAYTRAALALPWREGDIMLVNLRMAHGREPPTGERRVLVAMTSADS"
complement(2182..3918)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS12"
/protein_id="AAG02343.1"
/db_xref="GI:9937212"
/translation="MTLTLRDAFLDQAARTDAHAVVHGDVTVWYRELELRAGMART
LAAGAGTGLAVLRPEPPEPRAALLAVLTGAGYVPLADDDPPDRCHLDDCAA
LLAEHPSRDLTTPDEALAPAFDAAPRAGDPAYIYTSGSRPKPVLYVEQGA
LGYLAQARARYDGLSRLVHSSLFDMAYTSLMGVLSGGAHVLDAKAISQTP
PPAASARFLKVTVPSPHPLGLLPDSCLPQTGLVIGGALTSGALGPWRAHDPVT
VNEXPGETVCCCATVVRPGDVPDPAVPIGRPFAGTRLVLDADGPEVAVGGVGL
HIAGDQARGVLRPLTEERFPDPAAGSRMYRTGDLVREPRDGLVLRADGQ
VKVSGYRIPEGEIEAVLRGHAGKDCVAVGEADARLVAYVDPDPSPPGTAPAR
HAALGPYMPVATVPFELPPLTPNGKLDRLDLPAGPAGDAGDPTPAETLCELL
ARALGPEIDADADEFLTSGGTSITALKLVAGARRVGI RLELTTLRERTVRILAAQ
DAASPLAEGVPE"
complement(3915..4733)
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG02344.1"
/db_xref="GI:9937213"
/translation="MSKNAHWRSIRTGDAQGVVLAVDFYGTGRQENTFRHLCDLTD
PVEVHVPAPDGGSTATGAGLRWMTGTEGLVLAGRPVRLVGYCAGGVFASALA
DALVREGRHPRVLPNPSAGVATLTRDFGLGAMDLLTDGERAALLAETAIRRA
HADAPLVABRYAALYREGDLCERLGVDSAFCAELAAVLSYLAITLALDVPPT
PLWGAVLSLTSREHQGTDFTOVEHGFQVARAELLSSQVVAALLALLREHESR"
complement(4741..6024)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02345.1"
/db_xref="GI:9937214"
/translation="MTARVDTPTYLAVLAREAPILGSLARMSFVILPLALLIS
VRDATGSFVAGLISGALSATLTLFAPARAKLIDRRSGSLRVLTVYLGVLVIT
LAEEAPTAALLVAAVAGVFAFPPTMRVILHGRQPLRLTVYALDVSVEEV
FTVGPLAGGLIAVLAASMTVWLIATAGTACFVLSAATAAPASGEADEDRPHGR
PMALPGMRTIVLISGGVGLVQVLPFTIADHAGSPGAGGILLMSLSAGSAGVGLA
YGRJAWSTPVRVRVLTGFTLAVPLCLTASPVPAGAFALLVGLICLAPLFTTAYLL
VNDLVTASCTAPTEANTVSTANNNGFAAGSAAGVLLDSRGPTVTVTAFAVAATA
VMTVLRRTLLLGAGHPAAATPADRTAPAEAE"
complement(6021..6773)
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAG02346.1"
/db_xref="GI:9937215"
/translation="MDFELTRVNPWFSGCDGRPRVRLCALPYAGTAAVFKDWPAAL
PPGVELLTAHUPGRDGRTEPPPTATEAERLCEALPPSDLPVTVLHSGMALLGYE
EAARLAARGRAPNLLIAAACRPHPPDASGVTEAEALATLRWDPTALRDELM
EAVLPAVADITAGDGRHPRPRDLPDLPLKVIYIGADDDGTDWRTTLGWRACRACDEY
VVLPGCHVFLETDRAAVLTRVATDLAEAEVGA"
complement(6784..7797)
/codon_start=1
/transl_table=11
/product="clavaminic acid synthase-like protein"
/protein_id="AAG02347.1"
/db_xref="GI:9937216"
/translation="MTDPLRTVALTGEESAEDVLLRLTADVPVDSIVGLLHRLRLA
AQELPLRIARELTGMRLYDSPALVVTGVDDEIRGTPPAARPAPDEPRTDELLL
LLHAALGEAFGWATQNGRLVHDVLPVPGETAQMGSSSETELLHWHDEAFHPLRCD
YVGLLCRLNRHQAATVGMPLDLSRLTTEDRAVLLPEYLRIPDTSHTPAQNAATGRSA
ERFPAIAEMDDAPERVAVLFGDPEDPYLRIDPAYMSPAPDAAARAYDTVTALIDE
LRHVVDAGSLILLVDNYQAVHGRKPFAAAYDGRDRWLKRVNITRDLRSRSARRSATS
LLV"
complement(7859..9214)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02348.1"
/db_xref="GI:9937217"
/translation="MMKSSRLRDLQGGEDPVVAQESQDAGPTCQGGDGLNVAAL
AALLEVFPVRLPHHAGLRHVHSSQAAVALPQMVASATTGAGDGHQARRGO
VTGVGVRGAVGGDELGAERPHARQLDLGLLMAAERLADLPLQTVVQOGL
RGQVSHDGLGVDLPQGRGLLSLGRRRRGDGVGTAHASVGPQGPQPAAAEESCRG
LVAROODQALLRAVVEGPFORGEDAGVAAETVDHPDPVGEVGTGQVREOQGO
GCHVDREVPVAVGCGDDGVAGVGLRLAPVSAHVADPAHAGVLLGLLAVREOQGO
QRGGRAGVHCVPDLVQVEDLTHGQDRLLVADLLRQRTGLVDHRRPVPWPLARY
H"
complement(9211..17736)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS11-10"
/protein_id="AAG02349.1"
/db_xref="GI:9937218"
/translation="MASDALSAHLEVSDCRQRGITVTNDGSELAQNVAAVRRFERY
AIAPERTALLHKGATGYDELNRRAELTATRLADAGAPSTLVAVLPDPLDVATLC
ALLKGAACLPLDPCIAGRLREIMADASPDLVTRAVAPATGDPVPLFLDDAPT
CSAVLPHRSAGTASIAIYVLTPTDEKSENSVSYRDMARYLDDPTAGIPARAELR
LVAPLLSGRLVLDADETRPRPTREAPDMVEDVVAQVCAVGVDRVGRVDRFDFDL
GGKSLAAVQVVARLKKLLGVELPLRALFDAPTVEELAAVRAEQGGGVREEALEP
VGRSEPLPLSFAQQRLWFLDRLMPDRAFYTCMDAFVRGGIDLGALRALRMLVGRHE
TLRTAFVERDGVPIYOLVGPADGPGARRVAAPTRVDSLLEPAEREAEVNLVAAEART
PPRPADGALLRVVVARLADDDHVIYVSTHHIVSDAWSVGLVDELGRLYRCVTDGPA
ALPPAVQYADFAVQRAWAGPVQEEHLAYHKRALDCAVSVLRPLPMDHPRPAVOSER
GETVFALPDALVAALKLGRQGTALFEMTLGAFQVLLIARHAGQEDIVGVPAAGRT
RTEPELGVFFVNTLPLRAICAPGLSFRDLQVREAAALGAFADHQLDEALVEALAP
ERDLGNPLVQVTFQLLGTTPAARPLDGTVEVRYQVAVSQFDSLDTIKRADDSYR
GILNYCPDLDRRMEVLVGHVLTLLGAAAGDPRPIGELPLSDCAERLRLDDGFKR
DAAYAGCSVPERFAEVARTADARAVTCGATLTFEELNDRVERLAQALLGAGVTR
TPVAVLRPSTDSVALLAVMRAGGVYVPLDDWPDARTAYILDDTAAVITROLPA
LPGRLHVDPRPAADGLVPAPRIDPQAAIYIYTSGSTGAPKGVVRRSLNHLTSAL
QATFLGHDPYLAGADGVPPGDAKLTLTPTAFDTFDSMEQLSMGLAGHELFTVPEDVR
RDSALVRFVREHRIIDVITTSQLELLVSHGLLDGEWAPSMVWGVGASPSLMRTL
RDORTRCNLNGPTEAVDTACHDSDPADVPVIGTPLPHTHVLDLDRPRPVGV
AGEIYLGCTGLARGVYNPALTAREFVADPYDTPGSRLYETGDRARPPDGTLEYLG
RDDQIKLRGRFVEPGEIEAVLTHHPAVKEAAVDDAHARLYAVYTLAEGGAGPTDV
RRFAGKLPKRPVAVVLEALPLTSGNKDLRAKLPAPAGRGLVDFVFAVAPDMVE
EVVAGVCAVAGTVDRGVGDDDFELGGSLLVYQVMTIRIKLLGYEVLRELFDATY
BELAARAAATGELVGRGAAPPLGPDVRSGLPLTFAQORLWYLDQALPDSVSNMCD
AVRVGRDLDALRALTLVERHETLTAFAVERGVPHOVVSPADPAARAAAEVVR
IEAAGRTDEAVRDLVAAEARTPFPRADGALMERAVARLADDDHVLVTVTHIVSDGS
VDILVDELGRLYTHERVTDGPAGLPDVOYADFVQKSWMTGVPVREHLYAKWKALD
GAPSVLRPLPADHPRPAVOSQGETVEFFPLPVARLEALCREQCVTLFMALFGAFQV
LLARYSGDDVVVGVPTANTRAETEPLVGVFVNTLPVRVACSPELSFLALIDRVREA
```

ALCAFAHQDLPPEALVEALPBERDLGHHPVLQVTFOLLIDAPDERLVLHGTDCVSLGFG  
GVTSRDLSDVSVGRGRKCVLTICPDILFDRPRMEVLAGHYLDLALGAADDDPLRVG  
DPLSDSDVRLVLRRCVAAETPVAPCAETVPDFAFAVRATDAPALVHGDSLTFAE  
LDTRVLAALVRLRCVAAETPVAPCAETVPDFAFAVRATDAPALVHGDSLTFAE  
VAHLDLTAAPVITRDLPAOPGRVHLDRPADDDPLRLHRODAAV11FTSGST  
GAPKGVVVRHGSILYHLGHVRMAEGGPRNVAHTTAMTFDPSLEOFMLVAGHTLHV  
APEVRDRDEALVALVRAAIDVNLVTPSHLILLIEAGLLEDGRVPCTVLGGEAVPA  
ALWRTVRAETREFNLVGPTEATVDCHDSDPADVPVIGTUPHPTHVRVLDRL  
RVPVPGAEIYLGSTGLARGYLNRPALTAQRFVADPYDPTFGSLRYRTGDRWRPD  
GTLEYLGRYDDQIKIRGFRVEPEIEIENVLTHHPVAVKEAATVATDGAARLVALVPA  
PRPHGDSADQAOVEWMAVFEATHYDADADELTEFKWNDSPLDTPAPAEHMR  
ENVDTVARLLPAPAEVLEIGSTGLLMWRLLPHVTEYTGDESPVADMLRDRR  
RAHVRVLRHREATDGTGVRASDVLVNVSVVQYPPDRAYLDTVLARALDADRRR  
VFVQGNIALAPQFVARQALAHAGGAARDVARAAGEAFAMDEGLTWGDOVHDLASL  
ARSPVTVLEILPRGRHRNEMSLYRYDVLVHVGDRPAPAEVLEGLTWGDOVHDLASL  
SARLGRGGDALLVRGVANDRLTRONELLDAPARTAVTEDLWGLADSTPYRVSVW  
AAADPRGMDLLVRDADHDDGPLLVHPVPEPSAPLTNTPTPHPSAROGGSAADGLR  
SWLAERLPAHLLPARITEVDALPRGTGKLDRLGALGLVYAGRGARAGRPATAPRTG  
LERTLADARVRLGPEVGHENFVALGDSLLAVRAVARCARAGVRLTVRQLLSEQT  
VAALAAALEEESQ\*  
complement(18307..18660)  
/codon\_start=1  
/trans\_table=11

CDS

Query Match 5.6%; Score 112; DB 1; Length 77457;

Best Local Similarity 42.4%; Pred. No. 0.00038;  
Matches 794; Conservative 0; Mismatches 1065; Indels 13; Gaps 3;

QY 9 CCCCGACTTCCCTCGACGACACCGCTCCCTTACCGCGCGACCGCGCGCCGCGACTGCT 68  
DB 38086 CGCCAGGACCGCGCTGCAACACTGCTCCACCTCGCCCGCGGCTTCGCGCGCTGGA 38145  
QY 69 CTTAGCCTAGACGGGCGCTCTCTCGAGGTCTTCTGTTGGGACGGTGAAGCGACTGCGTC 128  
DB 38146 GGAGCGCACCCCGCGCGCTGCTGACCGTGAACACCGGTGCCACGACTGCTCGCGCA 38205  
QY 129 GAACCTGGTCTCTGGGGCGCGGTGTGACCGCGAGCCTCGAGACGCGACGGCCAGG 188  
DB 38206 CGACCTCGCCACCCCGAGCAGCCACCGCTCCCGGCGCGGCAAGGTGATCCCGCGGA 38265  
QY 189 AACCGTGCAGCTGACCGCATGACGTCGAGGCGCGCCAGCGATGCTGACCGCCCTGAACC 248  
DB 38266 GTACCGTGTGATCGCTGCACCGCTTGGACGTGGAGCGCGGCTGACCGCGGCGGT 38325  
QY 249 TGGCGCGGTCTTGGGTGAGAGGGTCTCCACCCCGACAGCTCTCTCTTACCGGTGCC 308  
DB 38326 GCGCGACTGATGCTCGGGAACCTCGCGCGCGCGCGAGACCACTGTCACCGCTCGCG 38385  
QY 309 GAACAGGTGGAGCTTCTCGCGGCCACCCGTCACAGAGAGAACCAATGACGCC 368  
DB 38386 CGGCGGACGCGGCTTACCGCCCTGCGCGGCTCCCGGCGCGGCAACG 38445  
QY 369 GGCATCTACGCGCGCGCTGCTCAGGAGCGCGCGCGGAGACATCGCCCTGATCTT 428  
DB 38446 CCGCGCGGTCCGCGCGCGCGCTTACCTGCTGCGGCGGCTCGGCGGATCGCCT 38505  
QY 429 CCGCGGTGCTGTCGCGCTGACGCGCGGATCCGCTTCGCGGTGCTGCTCGCTGCCGTCTA 488  
DB 38506 CCACCTCGCGAGTACC---TGGCGCGCGCGCGCGACACCGCTGCTCTACCCACCGCG 38562  
QY 489 CCACATGACCCCGCGCGGTGCTCTCGACACCCCGCGCGGTGTCACCCACCG 548  
DB 38563 GCGCTTCCCGCGCGCGGTGGAGCGGCTGCGCGGCGGACACCGGAGCGGCGCT 38622  
QY 549 CGCCTACCACTGACTACTGCACTCCGACCAAGAACAGCGCGCGCGGCTGGGACCA 608  
DB 38623 CPTCCGCGGCTCGGCTCTCTCCCGCGCGCGCGCGCGCGGCTGCTGCTCGCGCGGCG 38682  
QY 609 CCGGAGACACCGAGCGGTGCGCTTTCAGCACACCGACCGGACCGGTGATCGCGTGGGCG 668  
DB 38683 CTTACCGACACGAGCGGATGCGCGCTTCGCGGACGAGGTGGAACAGCGCCACGCGCC 38742  
QY 669 CGACTTCCCGGTGCTCGGCGGTGCTGCGGACCGCGGACCGCGGCTCGG 728

DB 38743 CGTCCGGGGGTGTGTCACGGCGCGGGGTGCCGACACCCCGCGCATGATCCAGCGTCG 38802  
QY 729 CGCGCGCGCGGTCTGCTCGCGTTCGCGCCAGCCAGCCAGCGCGCTCGCGAAATACCAAGA 788  
DB 38803 CGACCGAGCGGCGACGACGCGCGCGCTCGCGCGCAAACTGACCGCGCACCTCGTCTGGA 38862  
QY 789 GCAGTACTTACTGTGTCGACGACGCGGGGTTCACGTTTACCGCCCTCGCCGACCGCGT 848  
DB 38863 CGAGGTTCGCCCGACCGCGACCTCGACTTCTCTGCTGCTCTCGATCGGACCGT 38922  
QY 849 CATCGTCAACACCGACCGTTCGCGCGCGCACACCGCGCGCGAGATCAGAACGCGGAGTG 908  
DB 38923 GCTGCACAAGCTGAAGTTCGCGGAGGTTCGCTACGTGGCGGCGCAACGAGTTCTCTGACGC 38982  
QY 909 GTTCCCGGACCCCAAGATCCACTGGGACACACCGCGCGGAGAAATGGGTCTCGCTCATCGG 968  
DB 38983 CTATGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGGAGAACCTCTGCTGCTCGCTGACCGGA 39042  
QY 969 ACAGTCCGCTAGCAGCGCGGTCTTACACTCGCGGAACTCGCGGACCTGGGACCTTTCGCGG 1028  
DB 39043 CTGCGGGAGTGGGCGATGTGGCGCGCGCGCGCGCGCGCGCGCTGACCGCGCTACGCGAC 39102  
QY 1029 CAATTCGACTACCCGAAACACCGCTTCGCGGCGATCGAGTGGCGCGCGCGCGCTGTCGAGAT 1088  
DB 39103 CGCGCGCGACTGCGCTGCTACCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGATACGCGCGGAGA 39162  
QY 1089 CACCGAGACGAGCGGACGCGCACTGGGTGCTCGCGCGCGAGCATGGACGCTTACCGCAT 1148  
DB 39163 GGGGCTCGAGCTCTTCCCGCGGTGCTGCGCGCGGACACCGCGCGCGCGCTCATCGTCTC 39222  
QY 1149 CGGCTCCCCATGAGTACGCTTACTGGACAGGACCTTGGGACGCGGAGAGATTCAGCG 1208  
DB 39223 GGCGAGGACTCGACGAACCTCTCGCGCGCGGACGCGCGCTTACACACCGACGACCACT 39282  
QY 1209 CGAGGACTCACCCCGCAATGGGTGCTGAGTGGGCTGGGACTGTTACCGCGCGCTCACCTG 1268  
DB 39283 CGCGCGCTCGGCGACCTGAGGATGCGCGCGCGCGCGCGCTTCCGCGCGCGCGCGCG 39342  
QY 1269 GCCATCGAGCGCGCGCGGACCAAGCGCTCGCCATCGCGTGGTGAACAACTGAA 1328  
DB 39343 GTACGGGCGCGCGCACACGCGCGCGCGGATCGCGCGCTGTTACCGCGACCTGCT 39402  
QY 1329 GTAGCGGACGAGAGTCTCCCGACGAGCATCGGAGCGGTACAAACGCGGAGAGACTGAT 1388  
DB 39403 CGCGCTCGAAGACGTCGCGCTCGACGAGACTTCTTCGCGCTCGGCGGAGACTCGCTGCT 39462  
QY 1389 CGTCCGCGAGCTCGCGCTCGCGCGGACGCTGGGCGTGTATACCTCTCTGAGCACCCC 1448  
DB 39463 CGCGCTCGGCTGCTGTCGAGCTGCGGAGCGCTAGCGGTGGAGATCTCTGTCGCGCG 39522  
QY 1449 CGTGGGACGCTGACGAACCTACGTCACCGCGCACACACACCTCCCGCGCGGACCGTCA 1508  
DB 39523 CATGTTGACGAGCGCA-----CGGTGGCGGCTGGCGCGCGCGCGCGCGCG 39575  
QY 1509 CGGAGCGCGCTGCTGCGCATGGAACGCGGAGCATAGAGATCGAGCTCGACATCGCCTG 1568  
DB 39576 CGGAAGAGACCGCGCGCGGAGAGGTGTGCT---GTGACCAACGCGCGCGCTCACCGA 39632  
QY 1569 GGACCGCGGACGAACCTCGGCTCTCGGTGGCGCTTCCCGCGCGGAGAACCGCGGACAC 1628  
DB 39633 CTGTCTACCGAGCTCGCGCGCGGAGGTGACCTCAGCGCGGACGCGGACCGGCTGCA 39692  
QY 1629 GAACATCGGCAAGTACGAGGACGAGCTGTATGTCGACCGAGGACCTTCCGACCTTCGCGG 1688  
DB 39693 CTGCGCGCGCGCGCGGCGCGCTCACCAGAGCTCTCTCGCGACCATCTCGCGCGCGCG 39752  
QY 1689 GTACTGCTGCGCGCTTACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1748  
DB 39753 CGACGAACCTCTCG 39812  
QY 1749 GCACCTCGCGCTCTGTCGACACCGGAGCGTCTGAGGTCTTCTGTAACGCGCGGCGACAC 1808

Db 39813 GCCGCTGTCTTCGCCCGAGGAAACGGCTCTGGCTCTCTCCACAGTTCCACCCCGCAGACAG 39872  
Qy 1809 CGTGCTCTCCAGCAGTCCACTTCGCCGAGGGCGACACGGGAATCTCGCTCTACACCGA 1868  
Db 39873 CGCCTACAACATCCCCCTGCACATGCCCTGGCGGGCCCCCTGAACCCGGCCCGCCCTGGC 39932  
Qy 1869 CGCGGGCCCCGC 1880  
Db 39933 CGCGCCCTGGC 39944

Search completed: May 10, 2003, 08:28:52  
Job time : 6087.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 02:34:41 ; Search time 434.5 Seconds  
(without alignments)  
10371.113 Million cell updates/sec

Title: US-09-868-328B-2

Perfect score: 2001

Sequence: 1 gcgggcaccccgacttccc.....atgcacacgtctctgtgtt 2001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/qcgcdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 2001  | 100.0       | 2001   | 22 | Arthrobacter ureaf |
| 2          | 2000  | 100.0       | 2000   | 22 | Arthrobacter ureaf |
| 3          | 945.2 | 47.2        | 1752   | 24 | Levan fructotransf |
| 4          | 931.8 | 46.6        | 1551   | 20 | Arthrobacter nicot |
| 5          | 912   | 45.6        | 1467   | 24 | Levan fructotransf |
| 6          | 909.6 | 45.5        | 1452   | 20 | Arthrobacter nicot |
| 7          | 147.6 | 7.4         | 65140  | 22 | AAD17184           |
| 8          | 147.6 | 7.4         | 125401 | 22 | AAD17186           |
| 9          | 134.8 | 6.7         | 114955 | 20 | AA53491            |

|    |       |     |         |    |          |
|----|-------|-----|---------|----|----------|
| 10 | 119.8 | 6.0 | 3849    | 22 | AAF25795 |
| 11 | 112   | 5.6 | 5857    | 21 | AAA58471 |
| 12 | 109.2 | 5.5 | 27541   | 22 | AAD17185 |
| 13 | 101   | 5.0 | 4403765 | 22 | AA199683 |
| 14 | 99.8  | 5.0 | 114955  | 20 | AA53491  |
| 15 | 98.6  | 4.9 | 77536   | 21 | AAA14651 |
| 16 | 97.6  | 4.9 | 2712    | 20 | AAZ06825 |
| 17 | 97.6  | 4.9 | 2712    | 20 | AAH74538 |
| 18 | 94.8  | 4.7 | 4411529 | 22 | AA199682 |
| 19 | 93    | 4.6 | 5392    | 15 | AAQ64201 |
| 20 | 89.8  | 4.5 | 30690   | 21 | AAA92301 |
| 21 | 89.8  | 4.5 | 30690   | 22 | AAH79277 |
| 22 | 89.8  | 4.5 | 31422   | 21 | AAH92302 |
| 23 | 89.8  | 4.5 | 31422   | 22 | AAH79278 |
| 24 | 89.8  | 4.5 | 77536   | 21 | AAA14651 |
| 25 | 89.4  | 4.5 | 2291    | 9  | AA80309  |
| 26 | 88.6  | 4.4 | 2271    | 11 | AAQ06844 |
| 27 | 88.6  | 4.4 | 3391    | 22 | AAC86698 |
| 28 | 88.2  | 4.4 | 882     | 24 | ABK73444 |
| 29 | 85.6  | 4.3 | 44377   | 18 | AAT78508 |
| 30 | 85.6  | 4.3 | 44377   | 18 | AAT80414 |
| 31 | 84.4  | 4.2 | 2811    | 21 | AAC55794 |
| 32 | 84.4  | 4.2 | 18034   | 21 | AAC55841 |
| 33 | 84.2  | 4.2 | 985     | 24 | ABQ44816 |
| 34 | 84.2  | 4.2 | 985     | 24 | ABQ44817 |
| 35 | 84    | 4.2 | 4074    | 23 | AA576405 |
| 36 | 84    | 4.2 | 8438    | 15 | AAQ73500 |
| 37 | 83.2  | 4.2 | 11916   | 22 | AAH79279 |
| 38 | 83.2  | 4.2 | 24379   | 18 | AAH93095 |
| 39 | 83.2  | 4.2 | 24379   | 19 | AAV25925 |
| 40 | 83    | 4.1 | 1908    | 21 | AAC61403 |
| 41 | 83    | 4.1 | 4403765 | 22 | AA199683 |
| 42 | 82.2  | 4.1 | 2061    | 22 | AAF86409 |
| 43 | 81.8  | 4.1 | 1773    | 24 | ABK86017 |
| 44 | 81.8  | 4.1 | 1773    | 24 | ABK86018 |
| 45 | 81.4  | 4.1 | 1127    | 21 | AAA04777 |

ALIGNMENTS

RESULT 1

AAF30918

ID AAF30918 standard; DNA; 2001 BP.

XX

AC AAF30918;

XX

DT 09-JUL-2001 (first entry)

XX

DE Arthrobacter ureafaciens levan fructotransferase DNA.

XX

KW Levan fructotransferase; difructose dianhydride IV; sweetener; ds.

XX

OS Arthrobacter ureafaciens.

XX

Key Location/Qualifiers

FT CDS 361..1926

FT sig\_peptide 361..459

FT mat\_peptide 460..1923

FT sig\_peptide 361..459

FT mat\_peptide 460..1923

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Rhee S, Song K, Kim C, Ryu E, Lee Y;  
XX WPI: 2001-308483/32.  
DR P-PSDB: AAB82301.  
XX  
XX Producing difructose dianhydride IV from sucrose, involves reacting  
PT sugar solution in the presence of levansucrase to produce levan, and  
PT produce levan solution in the presence of levan fructotransferase to  
PT produce DFA IV.  
XX  
XX Claim 4: (Page 48; 72pp; English.  
XX  
XX The present sequence is that of *Arthrobacter ureafaciens* K2032 DNA  
CC encoding levan fructotransferase (see AAB82301). The DNA was  
CC obtained by PCR amplification of genomic DNA using degenerate  
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying  
CC the levan fructotransferase gene is claimed. Also claimed is  
CC *Escherichia coli* JUD81 (KCTC 0877BP), prepared by transforming  
CC *E. coli* DH5-alpha with pUDFA81. A claimed process for producing  
CC difructose dianhydride IV (DFA IV) from sucrose comprises  
CC subjecting sugar solution to reaction at room temperature or lower  
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase  
CC derived from *Zymomonas mobilis* to produce levan, purifying the  
CC levan from the reaction solution, and subjecting it to reaction at  
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in  
CC the presence of levan fructotransferase, preferably obtained by  
CC cultivation of *E. coli* JUD81. The product is useful as a  
CC low-calorie sweetener.  
XX  
XX Sequence, 2001 BP; 337 A; 813 C; 588 G; 263 T; 0 other;  
SQ

Query Match  
Best Local Similarity 100.0%; Score 2001; DB 22; Length 2001;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGTGACCCGACCTCCCTCGACGACACCGTCCCTACCGGCGACGCCCGGCC 60  
DB 1 GCGGTGACCCGACCTCCCTCGACGACACCGTCCCTACCGGCGACGCCCGGCC 60  
QY 61 CGACTGCTCTCAGCTAGACGGGCCCCCTCTCGAGGTCTTCGTGCGGGACGGTGAAGG 120  
DB 61 CGACTGCTCTCAGCTAGACGGGCCCCCTCTCGAGGTCTTCGTGCGGGACGGTGAAGG 120  
QY 121 ACTGCTGCAACCTGCTCTCTGCGGGGCGCGGCTGTGACCGCGAGCCTTCGAGACGGCA 180  
DB 121 ACTGCTGCAACCTGCTCTCTGCGGGGCGCGGCTGTGACCGCGAGCCTTCGAGACGGCA 180  
QY 181 CGGCCAGGAACCGTGACGTCACCGCATCGAGTGCAGGCGCCGCGCATGCTGACGCC 240  
DB 181 CGGCCAGGAACCGTGACGTCACCGCATCGAGTGCAGGCGCCGCGCATGCTGACGCC 240  
QY 241 CTGGAACCTGCGCCGCTTCTGGGCTCAGCAGGCTCCACCCGACAGCTCTCTTCTAC 300  
DB 241 CTGGAACCTGCGCCGCTTCTGGGCTCAGCAGGCTCCACCCGACAGCTCTCTTCTAC 300  
QY 301 CGCTGCCCCAAGCAGGGTGGAGCTTCTGTCGCGCCACCCGCTCCACGAGGAAACAGCA 360  
DB 301 CGCTGCCCCAAGCAGGGTGGAGCTTCTGTCGCGCCACCCGCTCCACGAGGAAACAGCA 360  
QY 361 ATGACCGCGGCATCTACGCGCGCGCTGCTCCAGGAGCGCGCGGAGCACTCGCC 420  
DB 361 ATGACCGCGGCATCTACGCGCGCGCTGCTCCAGGAGCGCGCGGAGCACTCGCC 420  
QY 421 CTGATCTTCGGGGTCTCTGCGGCTGCGAGCGCGGCGATCCGCTCGCGGCTCGCTCCGT 480  
DB 421 CTGATCTTCGGGGTCTCTGCGGCTGCGAGCGCGGCGATCCGCTCGCGGCTCGCTCCGT 480  
QY 481 GCGGTCTACCATGACGCCCCCGAGCGGCTGGCTTTCGAGACCCCGACCCCGGTACCC 540  
DB 481 GCGGTCTACCATGACGCCCCCGAGCGGCTGGCTTTCGAGACCCCGACCCCGGTACCC 540  
QY 541 ACCCAGGGCGCTACCACTGCTACTACCTGCACTCCGACCAAGAACACGGCCCCGGCGC 600

DB 541 ACCCAGGGCGCTACCACTGCTACTACCTGCACTCCGACCAAGAACACGGCCCCGGCGC 600  
QY 601 TGGGACACGCGAGCAGCAGCCGCGCTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 660  
DB 601 TGGGACACGCGAGCAGCAGCCGCGCTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 660  
QY 661 CTCGGCGCGAGCTTCCCGGCTGCTGCGGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCT 720  
DB 661 CTCGGCGCGAGCTTCCCGGCTGCTGCGGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCT 720  
QY 721 GGGTTCGGCGCGCGGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 780  
DB 721 GGGTTCGGCGCGCGGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 780  
QY 781 TACCAGGAGCAGTACCTCTACTGCTGACGACGCGGCGGCTTACGTTTACGCGGCTTGCCTC 840  
DB 781 TACCAGGAGCAGTACCTCTACTGCTGACGACGCGGCGGCTTACGTTTACGCGGCTTGCCTC 840  
QY 841 GACCCCGTTCATGCTCAACACCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 900  
DB 841 GACCCCGTTCATGCTCAACACCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 900  
QY 901 GCGGAGTGGTTCGGGACCGCCAGATCCACTGGGACACCGCCCGCGGAGATGGGTCTGC 960  
DB 901 GCGGAGTGGTTCGGGACCGCCAGATCCACTGGGACACCGCCCGCGGAGATGGGTCTGC 960  
QY 961 GTCATCGGAGCAGTGGGTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCT 1020  
DB 961 GTCATCGGAGCAGTGGGTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCT 1020  
QY 1021 CTTTGGCGCAACTTCGACTTACCGAACACCGCTTCCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1080  
DB 1021 CTTTGGCGCAACTTCGACTTACCGAACACCGCTTCCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1080  
QY 1081 TTCGAGATCAGCGACACGAGCGGACACCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1140  
DB 1081 TTCGAGATCAGCGACACGAGCGGACACCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1140  
QY 1141 TAGGGATCGGCTTCCCATGAGTACGCTTACTGACAGGACACCTGGGACGCGGAGCAG 1200  
DB 1141 TAGGGATCGGCTTCCCATGAGTACGCTTACTGACAGGACACCTGGGACGCGGAGCAG 1200  
QY 1201 TTCACGCGGACGACCTACCGCGCAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
DB 1201 TTCACGCGGAGACCTACCGCGCAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
QY 1261 GTCACCTGCGCATTCGATCGACGCGCGGACCAAGCGCTGCGCATCGCTGCTGCTGCTGCTGCTGCT 1320  
DB 1261 GTCACCTGCGCATTCGATCGACGCGCGGACCAAGCGCTGCGCATCGCTGCTGCTGCTGCTGCTGCT 1320  
QY 1321 AACTGGAAGTACGCGCAGCGGCTTCCCATGGAACGAGCGCATTCGACGCGCTTACACGCGGAC 1380  
DB 1321 AACTGGAAGTACGCGCAGCGGCTTCCCATGGAACGAGCGCATTCGACGCGCTTACACGCGGAC 1380  
QY 1381 AACTCGATCGTCGCGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1440  
DB 1381 AACTCGATCGTCGCGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1440  
QY 1441 AGCACCCGCTGGCAGCGCTGACGAACTACGTCACCGCACCACTTCCCGGACCGG 1500  
DB 1441 AGCACCCGCTGGCAGCGCTGACGAACTACGTCACCGCACCACTTCCCGGACCGG 1500  
QY 1501 ACCGCTGACGCGAGCGCTTCCCATGGAACGAGCGCATACGAGATCGAGCTCGAC 1560  
DB 1501 ACCGCTGACGCGAGCGCTTCCCATGGAACGAGCGCATACGAGATCGAGCTCGAC 1560  
QY 1561 ATCGCTGCGGACACCGCGAGCAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1620  
DB 1561 ATCGCTGCGGACACCGCGAGCAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1620  
QY 1621 CGGACACCAAGCATCGGCAAGTACGAGCAGACCTTCTAGCTGCGAGGAGCCTCCGAC 1680

Db 1621 CGGCACAGCATCGGCAAGTACGAGGACAGACCTGTACGTGACCGAGGACCCCTCCGAC 1680  
QY 1681 CTCGGCGGTACTCGTCCGCCCTACTCGGAGCGCGCGCCCATCGACCCCGCGGC 1740  
Db 1681 CTCGGCGGTACTCGTCCGCCCTACTCGGAGCGCGCGCCCATCGACCCCGCGGC 1740  
QY 1741 CGATCGGTGACCTCGGCATCTCGTCGACACCCAGAGCGTCTTCCTCAACGCC 1800  
Db 1741 CGATCGGTGACCTCGGCATCTCGTCGACACCCAGAGCGTCTTCCTCAACGCC 1800  
QY 1801 GGCACACGCTCTCCAGCAGCTCCACTTCGCGAGGCGACACGGATCTCGCTC 1860  
Db 1801 GGCACACGCTCTCCAGCAGCTCCACTTCGCGAGGCGACACGGATCTCGCTC 1860  
QY 1861 TACACCGAGCGCGCGCCGACACTTCACCGCATCTCGTCCGCGAGATTGGCCAGGCG 1920  
Db 1861 TACACCGAGCGCGCGCCGACACTTCACCGCATCTCGTCCGCGAGATTGGCCAGGCG 1920  
QY 1921 ATCTAGGCGATGACACACACCGCTTCACCGAAGCGCGCCCGGAGACGACGGCCGAC 1980  
Db 1921 ATCTAGGCGATGACACACACCGCTTCACCGAAGCGCGCGCCCGGAGACGACGGCCGAC 1980  
QY 1981 AATCGACACGCTCTCGTCGTT 2001  
Db 1981 AATCGACACGCTCTCGTCGTT 2001

RESULT 2

AAAF30919

ID AAF30919 standard; DNA; 2000 BP.

XX AC AAF30919;

XX DF 09-JUL-2001 (first entry)

XX DE Arthrobacter ureafaciens levan fructotransferase DNA.

XX KW Levan fructotransferase; difructose dianhydride IV; sweetener; ds.

XX OS Arthrobacter ureafaciens.

XX FH Key Location/Qualifiers

XX FT CDS 360..1925

XX FT sig\_peptide /\*tag= a

XX FT mat\_peptide /\*tag= b

XX FT /\*tag= b

XX WO200129185-A1.

XX PD 26-APR-2001.

XX PF 19-OCT-2000; 2000WO-KR01183.

XX PR 19-OCT-1999; 99KR-0045302.

XX PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

XX PA (REAL-) REALBIOTECH LTD.

XX PI Rhee S, Song K, Kim C, Ryu E, Lee Y;

XX WP1; 2001-308483/32.

XX PT Producing difructose dianhydride IV from sucrose, involves reacting

XX PT sugar solution in the presence of levansucrase to produce levan, and

XX PT reacting levan solution in the presence of levan fructotransferase to

XX PS produce DFA IV -

XX PS Disclosure; Page 8-9; 72pp; English.

XX CC The present sequence is that of Arthrobacter ureafaciens K2032 DNA

XX CC encoding levan fructotransferase (see AAB82301). The DNA was

CC obtained by PCR amplification of genomic DNA using degenerate  
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying  
CC the levan fructotransferase gene is claimed. Also claimed is  
CC Escherichia coli JUD81 (KCTC 0877BP), prepared by transforming  
CC E. coli DH5-alpha with pUDFA81. A claimed process for producing  
CC difructose dianhydride IV (DFA IV) from sucrose comprises  
CC subjecting sugar solution to reaction at room temperature or lower  
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase  
CC derived from Zymomonas mobilis to produce levan, purifying the  
CC levan from the reaction solution, and subjecting it to reaction at  
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in  
CC the presence of levan fructotransferase, preferably obtained by  
CC cultivation of E. coli JUD81. The product is useful as a  
CC low-calorie sweetener.  
XX  
SQ Sequence 2000 BP; 337 A; 813 C; 587 G; 263 T; 0 other;

Query Match 100.0%; Score 2000; DB 22; Length 2000;

Best Local Similarity 100.0%; Pred. No. 2.9e-310;

Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGGTGACACCCGACTTCCCTCGACGACACCGTCCCTACCGCGCGACCGCCGCC 61

Db 1 CGGTGACACCCGACTTCCCTCGACGACACCGTCCCTACCGCGCGACCGCCGCC 60

QY 62 GACTGCTCTCAGCCTAGACGGGCCCTCCTCGAGGTCTTCTCGGGGACGTGAGCGA 121

Db 61 GACTGCTCTCAGCCTAGACGGGCCCTCCTCGAGGTCTTCTCGGGGACGTGAGCGA 120

QY 122 CTGCGTCGAACCTGCTCCTCGGGGGCGGGGTGTGACCGCGGACCTTCGAGACGGCAC 181

Db 121 CTGCGTCGAACCTGCTCCTCGGGGGCGGGGTGTGACCGCGGACCTTCGAGACGGCAC 180

QY 182 GGCCAGGAACCGTGCACGTGACCGGATCGAGCGCGCCGACCGATGCTGAGCGCC 241

Db 181 GGCCAGGAACCGTGCACGTGACCGGATCGAGCGCGCCGACCGATGCTGAGCGCC 240

QY 242 CTGAACCTGCCCGGTTCTGGGTGACGAGCGCTCCACCCCGACAGCTCTCTCTTAC 301

Db 241 CTGAACCTGCCCGGTTCTGGGTGACGAGCGCTCCACCCCGACAGCTCTCTCTTAC 300

QY 302 GCTGCCGGAACAGGGTGGACGCTTCTCGGGCCGACCCGTCGACGAGGAAACAGCAA 361

Db 301 GCTGCCGGAACAGGGTGGACGCTTCTCGGGCCGACCCGTCGACGAGGAAACAGCAA 360

QY 362 TGAGCGCGCCATCTCACGCGCGCGTCTCAGGAGCGCGCGCGGACACTCGCC 421

Db 361 TGAGCGCGCCATCTCACGCGCGCGTCTCAGGAGCGCGCGCGGACACTCGCC 420

QY 422 TGATCTTCGGCGGTGCTGTCGCGCTGACGCGCGGCGCATCGGCTCGCTCCGTG 481

Db 421 TGATCTTCGGCGGTGCTGTCGCGCTGACGCGCGGCGCATCGGCTCGCTCCGTG 480

QY 482 CCGTCTACACATGACGCGCGCGCGGCTGCTGCGACCCCGACCGCGGTACCA 541

Db 481 CCGTCTACACATGACGCGCGCGCGGCTGCTGCGACCCCGACCGCGGTACCA 540

QY 542 CCCACGGCGCTTACGAGCTGTACTACTGCTCCGACGAGAAACAGCGCGCGCT 601

Db 541 CCCACGGCGCTTACGAGCTGTACTACTGCTCCGACGAGAAACAGCGCGCGCT 600

QY 602 GGGACACGCGAGCAGCAGCGCGGTTCCTTACGACACCGACCGCTGATGCCGC 661

Db 601 GGGACACGCGAGCAGCAGCGCGGTTCCTTACGACACCGACCGCTGATGCCGC 660

QY 662 TCGGGCGCGACTTCCCGTGTGGTTCGGGTCTCGGCGCGCGCGGACCGACAGGAG 721

Db 661 TCGGGCGCGACTTCCCGTGTGGTTCGGGTCTCGGCGCGCGCGGACCGACAGGAG 720

QY 722 GGTTCGGCGCGCGCGGTCTCGGCTCGGACCCAGCGCGCGCTCCGCAAGT 781

Db 721 GGTTCGGCGCGCGCGGTCTCGGCTCGGACCCAGCGCGCGCTCCGCAAGT 780

```
QY 782 ACCAGGAGCAGTACCTCTACTGTGTCGACGAGCGGGGTTACAGTTACACCGCCCTGCCCG 841
Db 781 ACCAGGAGCAGTACCTCTACTGTGTCGACGAGCGGGGTTACAGTTACACCGCCCTGCCCG 840
QY 842 ACCCGGTCATCGTCAACACACGAGGTGCGCGCCGACACGCGCGGAGATCGAGAACG 901
Db 841 ACCCGGTCATCGTCAACACACGAGGTGCGCGCCGACACGCGCGGAGATCGAGAACG 900
QY 902 CCGAGTGGTTCCGCGACCCCAAGATCCACTGGGACACCGCCCGGAGAAATGGTCTGCG 961
Db 901 CCGAGTGGTTCCGCGACCCCAAGATCCACTGGGACACCGCCCGGAGAAATGGTCTGCG 960
QY 962 TCATCGGACGACTGCGGTAGCGCGGTTCTACACCTCGCGCAACCTGCGGACTGGACAC 1021
Db 961 TCATCGGACGACTGCGGTAGCGCGGTTCTACACCTCGCGCAACCTGCGGACTGGACAC 1020
QY 1022 TTCGCGCAACTTCGACTACCCGCAACACGCGCTCGGCGGCATCGAGTGGCCGACCTGT 1081
Db 1021 TTCGCGCAACTTCGACTACCCGCAACACGCGCTCGGCGGCATCGAGTGGCCGACCTGT 1080
QY 1082 TCGAGATCACCGGACGAGCGGACACGCGCACTGGGTGCTCGCGCGCAGCATGGAGCGCT 1141
Db 1081 TCGAGATCACCGGACGAGCGGACACGCGCACTGGGTGCTCGCGCGCAGCATGGAGCGCT 1140
QY 1142 ACGGATCGGCTCCCGCATGACGTACGCTTACTGGACAGGCACTGGGACGCGGAGCAGT 1201
Db 1141 ACGGATCGGCTCCCGCATGACGTACGCTTACTGGACAGGCACTGGGACGCGGAGCAGT 1200
QY 1202 TCCACGCGGACGAGCTACACCCGCAATGGCTCGACTGGGCTGGAGTGGTACGCGGCGG 1261
Db 1201 TCCACGCGGACGAGCTACACCCGCAATGGCTCGACTGGGCTGGAGTGGTACGCGGCGG 1260
QY 1262 TCACCTGGCCATCGATCGAGCGCGCGGACCAAGCGCTCGGCCATCGCGTGGTGAACA 1321
Db 1261 TCACCTGGCCATCGATCGAGCGCGCGGACCAAGCGCTCGGCCATCGCGTGGTGAACA 1320
QY 1322 ACTGGAGTAGCGCGGACGAGCGTCCCGACGAGCTCCCGACGAGTCCGAGCGGACGAG 1381
Db 1321 ACTGGAGTAGCGCGGACGAGCGTCCCGACGAGCTCCCGACGAGTCCGAGCGGACGAG 1380
QY 1382 ACTCGATCGTCCGCGAGCTGCGGCTCGCGGACGAGCTGGGCTGGTACACCTCTCTGA 1441
Db 1381 ACTCGATCGTCCGCGAGCTGCGGCTCGCGGACGAGCTGGGCTGGTACACCTCTCTGA 1440
QY 1442 GCACCCCGTGGCAGCGCTGACGAAGTACGTACCGCCGACACACACTCCCGACCGGA 1501
Db 1441 GCACCCCGTGGCAGCGCTGACGAAGTACGTACCGCCGACACACACTCCCGACCGGA 1500
QY 1502 CCGTCGACGCGCGCGCTCTGCCATGGACGCGCGCATACGAGATCGAGCTCGACA 1561
Db 1501 CCGTCGACGCGCGCGCTCTGCCATGGACGCGCGCATACGAGATCGAGCTCGACA 1560
QY 1562 TCGCTGGGACACCGGACGAGCTGCGATCTCGGTGGGCGCTCCCGGACGGAACCC 1621
Db 1561 TCGCTGGGACACCGGACGAGCTGCGATCTCGGTGGGCGCTCCCGGACGGAACCC 1620
QY 1622 GGCACAGCAATCGGCAAGTACGAGAGCACTGTACGTTCGACCGAGGACCTCCCGAAC 1681
Db 1621 GGCACAGCAATCGGCAAGTACGAGAGCACTGTACGTTCGACCGAGGACCTCCCGAAC 1680
QY 1682 TCGCGGGTACTCGCTCGGCGCTTACTCGGAGCGCGCGCCCGCCATCGACCCCGGCGCC 1741
Db 1681 TCGCGGGTACTCGCTCGGCGCTTACTCGGAGCGCGCGCCCGCCATCGACCCCGGCGCC 1740
QY 1742 GATCGGTGACCTGCGCATCTCTGTCGACACCGACGAGCTGAGGCTTCGTCACGCGG 1801
Db 1741 GATCGGTGACCTGCGCATCTCTGTCGACACCGACGAGCTGAGGCTTCGTCACGCGG 1800
QY 1802 GGCACAGCTGCTCTCCGAGAGGTCCACTTCGCGGAGGCGGACGAGGAATCTCGCTCT 1861
Db 1801 GGCACAGCTGCTCTCCGAGAGGTCCACTTCGCGGAGGCGGACGAGGAATCTCGCTCT 1860
QY 1862 ACACCGACGGGCGCGCGGACACTTCACCGGATCTGCTCGCGGAGATTGGCCAGCGA 1921
```

```
Db 1861 ACACCGACGGGCGCGCGGACACTTCACCGGATCTGCTCGCGGAGATTGGCCAGCGA 1920
QY 1922 TCTAGCGGATGTCACACACACCGCTCACCAGCGCGCGCGGAGACGACGCGCGACA 1981
Db 1921 TCTAGCGGATGTCACACACACCGCTCACCAGCGCGCGCGGAGACGACGCGCGACA 1980
QY 1982 ATCGACACGTCCTCTCGTGT 2001
Db 1981 ATCGACACGTCCTCTCGTGT 2000

RESULT 3
ABL54835
ID ABL54835 standard; DNA; 1752 BP.
XX
AC ABL54835;
XX
XX 01-JUL-2002 (first entry)
XX
DE Levan fructotransferase related DNA sequence.
XX
XX Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;
KW levan; gene; ds.
XX
XX Arthrobacter nicotinovorans.
XX
XX Key Location/Qualifiers
XX CDS 4..1461
XX FT /*tag= a
XX FT /product= "levan fructotransferase related protein"
XX
XX JP2002017366-A.
XX
XX 22-JAN-2002.
XX
XX 06-JUL-2000; 2000JP-0205756.
XX
XX 06-JUL-2000; 2000JP-0205756.
XX
XX (NIPT ) NIPPON TENSEI SEITO KK.
XX (KAGA-) KAKAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI; 2002-287313/33.
XX
XX A new levan fructotransferase -
XX
XX Example 1; Fig 4; 15pp; Japanese.
XX
XX The invention relates to a new levan fructotransferase (LFTase). LFTase
XX decomposes the beta-2,6-fructoside bond of levan in polyfructan to
XX produce difructose dianhydride IV (DFA IV). The LFTase of the invention
XX is used for the preparation of DFA IV. The current sequence represents
XX a levan fructotransferase related DNA sequence.
XX
XX Sequence 1752 BP; 351 A; 592 C; 506 G; 303 T; 0 other;
XX
Query Match 47.2%; Score 945.2; DB 24; Length 1752;
Best Local Similarity 74.3%; Pred. No. 2.9e-142;
Matches 1237; Conservative 0; Mismatches 413; Indels 14; Gaps 3;
QY 314 AGGTGGAGCGTTCGTGCGGCGCCACCGCTCCAGGAGAGAAC--AGCAATGACGCGCG 371
Db 42 AGCTTTGGGTGTCGTGCGGCGCGCCGCGCAACTGAGAGGAAACGAATCGATGACGTATGA 101
QY 372 CATCTCACGCGCGCGCTGCTCCAGGAGCGCGCGGAGCACTCGCCTGATCTTCGG 431
Db 102 CATCTTCGCGCGCACTGCCCTGCAAGGTGACGGGTTGGTGGCTTTCATGAG 161
QY 432 CGGTGCTGTGCGCGCTGCGAGCGCGGCACTCGGCTCGGCGCTCGCTCGCTTACCA 491
Db 162 CAATGCCATTCCGTTGCGCGCGCA-----CGCCAGGCACTCCCTCCGCGCATCTACCA 215
```



```
Qy 492 CATGACGCCCCAGCGGCTGGCTCTGCGACACCCCAACGCGCGGTACACACCCAGCGGC 551
Db 216 CATGACCCCGGCTCGGGCTGCTATGTGATCCGAGGAGACCCGTATACATAAAGCGGC 275
Qy 552 CTACGAGCTGTACTACCTGCACTCCGACACAGAACAGCGCCCGGCTGGGACACGC 611
Db 276 CTACGAGCTGTACTACCTGCACTCCGCGCAGAACAGCGCGGGATGGGACACGC 335
Qy 612 GAGCAGCACCGGCGCTGGCTTTCACGCACACAGCGCACCGTATGCCGCTGCGGCCGA 671
Db 336 GACCACCGCGGACGGAGTCTTACACCCACCATGAGTGGTGATGCCAATGCACCCGA 395
Qy 672 CTTCCCGCTGTGGCTCGGGTGGCGGTCTGGACACCGGACCGGAGGTTTCGGCGC 731
Db 396 CTTCCCGCTGTGGTGGGATGGCAGTAGTGACACCGCCCAACACCGCGCGCTTCGGCGC 455
Qy 732 CGGCGCGTGTGCGGCTGCGGACCGACCGGACCGGCGCTCCGCAAGTACCAGGACA 791
Db 456 CGGCGCAGTATCGGCTGCGGACCCCAACCCACCGACG-----GAAATTCAGGAACA 509
Qy 792 GTACCTTCTACTGGTGACCGGAGCGGGGTTTACGTTTACCGGCTTCCCGGACCCCGTCA 851
Db 510 GTACCTTCTACTGGTCCACGGATGGCGGTACTCTTTCACCGCATTTGCCCTGACCCGCTAT 569
Qy 852 CGTCAACACCGAGGTGCGCGCGGACACCGCGCGCGGATCGAGATCGAGAACCGCGAGTGGTT 911
Db 570 TGTGAACACTGTGAGCAGGCGGCGGACCCACCGCGCGGAGTGGAGAACCGCAGATGGTT 629
Qy 912 CCGGACCCCAAGATCCACTGGGACACCGCGCGGGAATGGTCTCGCTCATCGGACG 971
Db 630 CCGGACCCCAAGATCCACTGGGACCGGACCGGCAACGAGTGGGTCTGTGTCATCGGACG 689
Qy 972 ACTGGGTACGCGCGGTTTCTACACCTCGCGCAACCTGCGGACCTGGACACTTCGCGCGCA 1031
Db 690 GCGCGCTACGCTGCTTCTACACCTCTCCCACTCGGGATGGCAATGGAATCCAA 749
Qy 1032 CTTGACTACCGGAACACCGCTCGGGGATCGAGTGGCCCGGACCTGTTCGAGATCAC 1091
Db 750 CTTGACTACCGGAACACCGCTCGGGGATCGAGTGGCCCGGATCTGTTCGAATGAC 809
Qy 1092 CGCAGACGCGGACACCGGCTGCTGCGCGGACGAGTGGACGCTACGCGATCGG 1151
Db 810 CGCAGGACGGAACCGCGGCTGCTGCGGCGGACGAGTGGACGCTACAGCATCGG 869
Qy 1152 CTTCCCATGAGTACGCTACTGACAGGACCTGCGGACGCGGAGCAGTTCCACCGCA 1211
Db 870 CTTGCCCATGAGCTTGGCTACTGACAGGTTCTATGGAACGCGACAGCATTCATCGCGCA 929
Qy 1212 CGACCTACCCCGCAATGGCTGCACTGGGGTGGGACTGGTACGCGCGCGTCACTCGGCC 1271
Db 930 CAACCTACACACAGTGGCTTGAATGGGATGGGACTGGTACGCGCGCTGACCTGGCC 989
Qy 1272 ATCGATCAGCGCGCGGACGACCAAGCGCTCGCCATCGCTGGATGACAACTGGAAGTA 1331
Db 990 GCGCGTGGAGACGCTGAGACCAAGCGGCTTGGCACACGCTGGATGAACAACTGGAATA 1049
Qy 1332 CGCGCAGCGGCTCCCGACCGCATCCGAGCGCTACAGGCGGAGAGTGCATCGT 1391
Db 1050 TGCGCGCGCAAGCTGCGCCACGAGCGCTCGATGGCTTATACGGGCAAAATTCATCAT 1109
Qy 1392 CCGGAGCTGCGGCTCGCGGACAGCTGGGGCTGTGTACACCTCTCTGAGCACCCCGT 1451
Db 1110 GCGGAGCTCAGGCTCGAGGCGCAATCGGGCGGCTGTGTACACCTTGTCTCAGCAGCCCGT 1169
Qy 1452 GGCAGCGTGAAGAACTAGTCAACCGGACCAACACACACTCCCGGACCGGCTCGAGCG 1511
Db 1170 TCCGGCGCTTTCGAACACTATGCACTCCAGCAGCAACCCCTTCGGGACCGCAGTCAACGG 1229
Qy 1512 CAGCGCGCTCTGCTGCAAGGACGCGCATACGAGATCGAGTCTGAGCTCGCTCGGTA 1571
Db 1230 CAGTTTCTGCTGCTGAGGCGGCGGCTGTGAACTGGAACCTGATATTTATGGA 1289
Qy 1572 CACCGCAGCAAGCTCGGCATCTCGGTGGCGGCTTCCCGGACGGAACCCGCGCACAGAA 1631
```

```
Db 1290 CACGCGACGCAACGTTGGAGTCTCGGTGGCGGCTCGTCCGATGGCAGCGCCATACGAA 1349
Qy 1632 CATCGGCAAGTACGAGCAGACCTGTACGTGACGAGGACCTCCGACCTCGCGGGTA 1691
Db 1350 CATCGGCAAAATACGCTGACGAGTGTAGTGCATCGCATCTCGGAGCAAGCGGTA 1409
Qy 1692 CTCGCTCGCCCTACTCTCGGAGCGCGCGCCCATGACCCCGCGCGGATCGTGA 1751
Db 1410 TCGCTGCGACCTACACCGCGCGCGCCCATGATGCGAACGCGAGATCGGTCCA 1469
Qy 1752 CTTGGGATCTCTGTCGACACCCAGCAGGCTCGAGTCTTCTGTCACCGCGGCGCACCGT 1811
Db 1470 CTTGGGATCTCTGTCGACACCCAAAGTGTGAGTGTTCGTAATTCGGGCGCACCGT 1529
Qy 1812 GCTCTCCACGAGTCTTCTGCGGAGGCGGACACGGAATCTCGTCTACACGACGG 1871
Db 1530 GCTTTCGACAGGTCGACTTCGCGCGCGGACACGCGGATCTCCCTCTATGCGGACGG 1589
Qy 1872 CGGCGCGGACACTTCACCGGATCGTCTCGCGGATTTGGCCAGGCGATCTAGGCGAT 1931
Db 1590 CGGTCCGCGCAACTTCACCGGATCACCATCGCGAGTTGCGGAACCCCATCTAAGCGCTG 1649
Qy 1932 GCACACACACCGCTCACCGAAGCGCGCGCGGCGGAGGACGG 1975
Db 1650 CTTCCCGCGCTGGAAGGACGACGCGCGCGCTGCGAGCGCG 1693
```

## RESULT 4

AA19827 standard; DNA; 1551 BP.

AC AAX19827;

DT 10-JUN-1999 (first entry)

DE Arthrobacter nicotinovorans levan fructotransferase encoding DNA #2.

KW Arthrobacter nicotinovorans; levan fructotransferase; ds.

OS Arthrobacter nicotinovorans.

PN JP11069978-A.

PD 16-MAR-1999.

PF 28-AUG-1997; 97JP-0232421.

PR 28-AUG-1997; 97JP-0232421.

PA (NIOC ) NIPPON OIL CO LTD.

DR WPI; 1999-247463/21.

P-PSDB; AAY04105.

PT Levan fructotransferase gene - for recombinant production of levan fructotransferase

PS Claim 4; Page 10-11; 14pp; Japanese.

CC The present sequence encodes Arthrobacter nicotinovorans levan fructotransferase. The present invention also describes a method for the preparation of levan fructotransferase in which a transformant is cultured in a medium and levan fructotransferase is collected from the culture. The method can prepare levan fructotransferase in a large amount.

SQ Sequence 1551 BP; 315 A; 524 C; 443 G; 269 T; 0 other;

Query Match 46.6%; Score 931.8; DB 20; Length 1551;

Best Local Similarity 75.8%; Pred. No. 4e-140; Mismatches 367; Indels 12; Gaps 2;

Matches 1184; Conservative 0;

QY 361 ATGACGCGGCCATCTACGCGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 420  
Db 1 ATGACGCGGCCATCTCTCGCGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 60  
QY 421 CTGATCTTGGCGGCTGTGCGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 480  
Db 61 CTGATCTTGGCGGCTGTGCGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 114  
QY 481 GCGGCTTACACATGACGCGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 540  
Db 115 GCGATCTACACATGACGCGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 174  
QY 541 ACCGAGCGCGCTTACGAGTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
Db 175 ACAAGCGCGCTTACGAGTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 234  
QY 601 TGGGACACGCGAGCAGCAGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 660  
Db 235 TGGGACACGCGAGCAGCAGCGCGTCTCCAGGAGCGCGCGGAGCATCGCC 294  
QY 661 CTCGCGCGCGCTTCCCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Db 295 ATCAACCGCGCTTCCCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 354  
QY 721 GGGTTCGCGCGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Db 355 GGGTTCGCGCGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 408  
QY 781 TACGAGGAGCAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Db 409 TCCAGGAGCAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 468  
QY 841 GACCGCGCTCATGCTCAACACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Db 469 GACCGCGCTCATGCTCAACACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528  
QY 901 GCGGAGTGTGCTGCGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 529 GCAGATGTGCTGCGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 588  
QY 961 GTCATCGGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
Db 589 GTCATCGGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 648  
QY 1021 CTTGCGCGCACTTGCATACCGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1080  
Db 649 TGAAGTCAACTTGCATACCGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 708  
QY 1081 TTCGAGATCACCGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
Db 709 TTCGAGATCACCGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 768  
QY 1141 TAGGCGATCGGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
Db 769 TAGAGATCGGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 828  
QY 1201 TTCGAGCGCGAGCAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
Db 829 TTCGAGCGCGAGCAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888  
QY 1261 GTCACCTGGCCATCGATGAGCGCGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1320  
Db 889 GTGAGTGGCGCGCGTGGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 948  
QY 1321 AACTGGAAGTACCGCGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
Db 949 AACTGGAAGTACCGCGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008  
QY 1381 AACTCGATCGTCCGCGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
Db 1009 AACTCGATCGTCCGCGAGCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068  
QY 1441 ACCACCGCGTGGCAGCGCTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500

Db 1069 AGCACCGCGTTCGCGCGTTCGAACTATGCCACCTCCAGCACCCCTTCGCGACCGC 1128  
QY 1501 ACCGTCAGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
Db 1129 ACAGTCAACGCGAGTTCGTAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1188  
QY 1561 ATGCGCTGGGACACCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620  
Db 1189 ATTTCATGGGACACCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1248  
QY 1621 CGGCACACGAGCATCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680  
Db 1249 CGGCATACGAGCATCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1308  
QY 1681 CTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
Db 1309 CAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1368  
QY 1741 CGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
Db 1369 AGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1428  
QY 1801 GGCACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860  
Db 1429 GGCACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1488  
QY 1861 TACACCGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920  
Db 1489 TATCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1548  
QY 1921 ATC 1923  
Db 1549 ATC 1551

## RESULT 5

ABL54834  
ID ABL54834 standard; DNA; 1467 BP.  
XX AC ABL54834;  
XX AC ABL54834;  
DT 01-JUL-2002 (first entry)  
XX  
DE Levan fructotransferase encoding sequence.  
XX  
KW Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;  
KW levan; gene; ds.  
XX Arthrobacter nicotinovorans.  
XX  
FH Key Location/Qualifiers  
FT CDS 4..1461  
FT /\*tag= a  
FT /product= "levan fructotransferase"  
XX  
PN JP2002017366-A.  
XX  
PD 22-JAN-2002.  
XX  
PF 06-JUL-2000; 2000JP-0205756.  
XX  
PR 06-JUL-2000; 2000JP-0205756.  
XX  
PA (NIP) NIPPON TENSAN SEITO KK.  
PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.  
DR WPI; 2002-287313/33.  
DR P-PSDB; ABB10073.  
PT A new levan fructotransferase  
XX  
PS Claim 3; Fig 2; 15pp; Japanese.



CC The present sequence encodes *Arthrobacter nicotinovorans* levan  
CC fructotransferase. The present invention also describes a method  
CC for the preparation of levan fructotransferase in which a transformant  
CC is cultured in a medium and levan fructotransferase is collected from  
CC the culture. The method can prepare levan fructotransferase in a  
CC large amount.  
XX  
SO

Sequence: 1452 BP; 299 A; 491 C; 416 G; 246 T; 0 other;

Query Match: 45.5%; Score 909.6; DB 20; Length 1452;  
Best Local Similarity 77.0%; Pred. No. 1.4e-136;  
Matches 1123; Conservative 0; Mismatches 329; Indels 6; Gaps 1;

```
QY 466 CCGGGCTCGCTCGTGGCGTCTACCAATGACGCCCCCGGCTGGCTGGGACCCC 525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CAGGCATCCCTCGGGGCGATCTACCAATGACGCCCCCGGCTGGCTGGTATGTGATCG 60

QY 526 CAACGCCGCGTCAACACCCAGCGGCTTACAGCTGTACTACCTGCACTCCGACCAAC 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CAGGACCCCGTACATAAAGCGGCTTACAGCTGTACTACCTGCACTCCGACCAAC 120

QY 586 AAGCGCCCGCGGCTGGGACCGGAGACGACGCGGCTGGCTTTCAGCGCACAC 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 AAGCGACCGGCGGATGGGACCGGAGACGACGCGGAGTGTCTTACACCCACCAT 180

QY 646 GGACCGTGATCGGCTGGCGGCGGACTTCCCGTGTGGTTCGGGTCGGCGTGGC 705
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GGAAGTGGTGATGCAATGCAACCGGACTTCCCGTGTGGTTCGGGATCGGAGTAGTGGAC 240

QY 706 ACCGGAACACGCGAGGTTTCGGCGCGGCGGCTGCTGCGCTCGCGACCCAGCGGAC 765
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ACGCCAACACCGCGGCTTCGGCGCGGCGGAGTCACTGCGCTCGCGACCCAGCGGAC 300

QY 766 GACGGGCTCCGCAAGTACCAGGACGAGTACTCTACTGTCGACCGAGCGGCTTCAACG 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GAGC-----GAAATTCAGGAACAGTACTTTACTTGGTCCAGGATGCGGCTACTCC 354

QY 826 TTACCGCCCTCGCCGACCCGTCATCGTCAACACCGAGCGTTCGCGCGGACCGCC 885
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 TTACCGCGATTTGCTGACCGCGGCTATTGTGAACACTGATGGAGCGGACCGGACCC 414

QY 886 GCGGAGATCAGACGCGGAGTGTTCGCGGACCCCAAGATCCACTGGGACCGCGCCG 945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 GCGGAGTGGAGAACGCGAGAAATGTTCCGCGACCCGAAATTCATCGGGAGCGACGCG 474

QY 946 GGAAGTGGGTCTCGGTATCGGACGACTGCGGTACGCGGCTTCTACACTCGCGAAC 1005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 AAGAGTGGGTCTGTGTCATCGGACGCGGCGCTACGCTGCTTCTACACTCTCCCAAC 534

QY 1006 CTGCGGACTGGACACTTCCCGCACTTCGACTACCGAACCGCCCTCGGGGGATC 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 CTGCGGATTTGGAAATGGAAGTCCAACTTCGACTACCCCAACCGCCCTCGGGGTATC 594

QY 1066 GAGTGGCCCGACCTGTTCGAGATCAGCGAGACGCGGAGCGGACGCTGGGTCTCGCC 1125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 GAATTCGCGGATCTGTTCGAAATGACCGAGAGAGGAAACCGCGGACTGGGTTCGGG 654

QY 1126 GCCAGCATGACGCGCTACGCGATCGGCGCTCCCATGACGTACGCTTCTGACAGCAGC 1185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 GCGAGCATGGACGCTTACAGCATCGGCTTGCCCATGACCTTTCCCTTCTGACAGGTTCA 714

QY 1186 TGGGAGCGGAGAGTTCACGCGGAGGACCTCACCCCGGAATGGCTTCGACTGGGCTGG 1245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 TGAAGCGGACACGATTCATTCGCGGACAACTTCACACCAAGTGGCTTGAATGGGATGG 774

QY 1246 GACTGTGACGCGCGCTGACCTGCGCATGATCGACGCGCGGAGACCAAGCGCTTCGCC 1305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 775 GACTGTGACGCGCGCTGACCTGCGCGGCGGTGGAAGACACTGAGACCAAGCGCTTGGC 834

QY 1306 ATCCGCTGGATGAACAACTGGAAGTACGCGGACGAGCTCCCGACCGACGATCCGAC 1365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 835 ACAGCGTGGATGAACAACTGGAATATGCGCGCGCAAGCTGTCGCCAGCGGCGTCCGAT 894
```

```
QY 1366 GGCTAACACGGGAGAACTCGATCGTCCGGAGAGTCCGGCTCGCCACAGCCTGGCGGC 1425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 895 GGCTATAACGGGCAAAATTCATCATCAGCGGAGTCAAGCTCGAGCGCCCAATCGGGCGC 954

QY 1426 TGGTACACCTCTCTGAGCACCCCGTGGCAGCGCTGACGAACCTACGTCACCGCACCA 1485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 TGGTACACCTTGTCTACGACGCGCGTTCGCGGCGTTCGAACTATGCACTCCAGCACC 1014

QY 1486 ACATCTCCCGACCGGACCGTTCGAGCGGAGCGCGCTCTGCTGTCATGGAACGACGCG 1545
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 ACCTTCGCGACCGCACAGTCAACGGCAGTTTCGTACTTCCGTGGAGCGCGGCGGTAT 1074

QY 1546 GAGATCAGCTCGACATCGCTGGGACACCGCGACGAGCTGTCAGTCTCGAC 1605
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1075 GAACTGGAACCTCGATATTTTCATGGACACGCGACGCAACGTTGGGAGTCTCGGTGG 1134

QY 1606 TCCCGGACGGAACCGGCGCACGACATCGGCAAGTACGAGACGAGTCTGACGTCGAC 1665
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1135 TCGTCCGATGGACCGCCCATACGAACATCGGCAAACTCGGTGACGAGTTGTAGTCTCAT 1194

QY 1666 CGAGGACCTTCGACCTCGCGGCTACTCGCTCGCCCGCTACTCGCGAGCGCGCCGCC 1725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1195 CGCGCATCTTCGGAGCAAGCGTTATGCTGCGTGGCACCCCTACACCGCGCGCGCGCC 1254

QY 1726 ATGACCGCGCGCGCGGATCCGTCGACCTGCGGATCTCTGTCGACACCGAGAGCTGAG 1785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1255 ATGACGCAACCGAGATCCGTCACCTCGCATCTTTGTAGACACCCCAAAAGTGTGAG 1314

QY 1786 GTCTTCGTCACCGCGCGCACACGCTGCTCTCCAGCAGGTCCACTTCGCGGAGGCGAC 1845
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1315 GTGTTGTAATTCGCGGACACGCGTGTTCGACGAGGTGCACTTCGCGCGCGGCGAC 1374

QY 1846 ACGGGAATCTGCTTACACCGAGCGCGCGCCGACACCTTCACCGGCATCTGTCGTCGC 1905
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1375 ACGGGATCTCCCTCTATGCGGACGCGCGTCCGCGCAACTTCACCGGCATCACCATCCG 1434

QY 1906 GAGATTGGCCAGGGATC 1923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1435 GAGTTGGGAAACCCCATC 1452

RESULT 7
AAD17184
ID AAD17184 standard; DNA; 65140 BP.
XX
AC AAD17184;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nysI; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
CDS complement (1..1035)
   /tag= a
   /product= "NysD2 partial protein"
   /note= "CDS does not include stop codon"
CDS complement (1036..2576)
   /tag= b
   /product= "NysD1 protein"
   /tag= c
   /product= "NysA protein"
   /tag= d
   /product= "NysB protein"
   /tag= e
   /product= "NysC protein"
```



```
QY 1088 TCACCGGACGACGAGGACAGCCAGCAGTGGTGTCTGGCGCCAGCATGGAGCCCTACGGCA 1147
Db 20859 ACATCGCGCTGCGAGCTGCGGACCGCGCCCTCGCGACCTCTCTCGGACCGTCC 20918
QY 1148 TCGGCTCCCATACGCTAGCCTACTGAGACAGGCACTGGGAGGCGGACGATTCACG 1207
Db 20919 CGGCGGGACCGCGCTACCGCTGCTGCACACCGCGCGCTGCTGCGAGCGGCTCC 20978
QY 1208 CCAGCAGACCTACCCCGCATGGCTGACCTGGGCTGGGACTGGTACGCGCGCTGACCT 1267
Db 20979 TCGGCTCCCTACCGCGCAGCGCTGACACACCTGCTCGGCGCCAGCGCGCGCT 21038
QY 1268 GGCATCGATCGAGCGCGCGGACGACCAAGCGCTCGCCATCGGCTGGATGAACAACCTGA 1327
Db 21039 GGCACCTGACGAGCGCGCGGACCGCGCTGACCTGAGCGCTTCTGCTCTCTGCTCG 21098
QY 1328 AGTACGCGCGAGGAGTCTCCACGACGACGCTACAGCGCTACAGGGCGCAACTCGA 1387
Db 21099 TCGCGCGCGCGCTCGGACGCGCGGACAGCGCAACTACGCGCGCGCAAGCGCTTCTG 21158
QY 1388 TCGTCCGCGAGCTGCGGCTGCG--CGACAGCTGGCGGTGTACACCTCTCTGAGCAC 1445
Db 21159 ACGCCCTCGCGCGCGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCTCTCTGCGCG 21218
QY 1446 CCCCGTGGAGCGCTGACGAACCTACGTCACCGCGCGCGCGCGCGCTGCGCGCGCG 1505
Db 21219 CGTGGACCGCAGAGCGTGGCGATGAC--AGCAGCGCTGTCGAGCTCGACGTCGAGCG 21276
QY 1506 CGACGCGAGCGCGCTCTGCGCATGGAGCGGACGCGCATAGGAGATCGAGCTGACATCG 1565
Db 21277 CGCCCGCTCGGCGATGCGCGCGCTGACCTTGSAAACGAGCGACCGCGCTCTTCGACG 21336
QY 1566 CTGGGACACGCGACGAACCTCGGCTATCTGCGTGGCGCGCTCC--CCGACGGAACCG 1622
Db 21337 CTGCGCGCGCGCGCGCGCGCGCTGCGCGCGCTGATCGCGACCGCGCGCGCGCG 21456
QY 1683 CGCGGGTACTGCTGCGCGCGCTTCTGCGGAGCG-----GCGCGCGCGCTGACCGCG 1736
Db 21457 CGCGCGAGTCTGCGACGCGCGCGCGCGCTGCGCGCGCGCGCGCGCTGCGCGCGCG 21516
QY 1737 CGCGCGATCGTGCACCTGCGCATCTCTGTCGACACCGCGAGCGTCTGAGTCTTCTCAA 1796
Db 21517 CGCGCGCGGAGCGCTCTGGAATCTGTCGACCGCGAGATCGCGCGCGCTCTCGGCG 21576
QY 1797 CGCGCGCGACCGCTGCTCTCCAGCAGGTCTTCTGCGCGCGCGCGCGCGCGCGATCTC 1856
Db 21577 CGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21636
QY 1857 GCTCTACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1904
Db 21637 CACCGCGTGAACCTCGGCAACCGCGCTGACACCGCGCGCGCGCGCGCGCG 21684
```

## RESULT 8

AD17186  
ID AD17186 standard; DNA; 125401 BP;

XX AC AD17186;

XX DT 29-NOV-2001 (first entry)

DE Streptomyces noursei nystatin PKS gene cluster DNA.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; ds.

OS Streptomyces noursei.

XX Key Location/Qualifiers

```
FT CDS 6337..34771  
FT /*tag= a  
FT /product= "NysI complete protein"  
FT 34792..51099  
FT /*tag= b  
FT /product= "NysJ protein"  
FT 51155..57355  
FT /*tag= c  
FT /product= "NysK protein"  
FT 57503..58687  
FT /*tag= d  
FT /product= "NysL protein"  
FT complement (58786..58980)  
FT /*tag= e  
FT /product= "NysM protein"  
FT /note= "CDS does not include start codon"  
FT /*tag= f  
FT /product= "NysN protein"  
FT /note= "CDS does not include start codon"  
FT complement (60238..61296)  
FT /*tag= g  
FT /product= "NysD2 complete protein"  
FT 120628..121308  
FT /*tag= h  
FT /product= "NysR4 (long) protein"  
XX  
PN W0200159126-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 08-FEB-2001; 2001WO-GB00509.  
XX  
PR 08-FEB-2000; 2000GB-0002840.  
PR 10-APR-2000; 2000GB-0008786.  
PR 14-APR-2000; 2000GB-0009387.  
XX  
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.  
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.  
PA (ALPH-) ALPHARMA AS.  
PA (SINV-) SINVENT AS.  
PA (DZIE-) DZIEGLEWSKA H.  
PA (ZOTC/) ZOTCHEV S B.  
PA (SERU/) SEKUROVA O N.  
PA (FJAE/) FJAERVIK E.  
PA (BRAU/) BRAUTASET T.  
PA (STRO/) STROM A R.  
XX  
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;  
XX  
XX WPI: 2001-557614/62.  
DR P-PSDB; AAEL10143, AAEL10144, AAEL10145, AAEL10146, AAEL10147, AAEL10148,  
DR AAEL10149, AAEL10150.  
XX  
XX New nystatin polyketide synthase polynucleotides and polypeptides,  
XX useful as antibiotics and antifungals -  
XX  
XX Claim 1: Page 188-254; 266pp; English.  
PS  
XX  
XX The present invention relates to the cloning and sequencing of the gene  
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme  
XX involved in the biosynthesis of the macrolide antibiotic nystatin.  
XX The nystatin PKS is useful as antifungal antibiotics. The present  
XX sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.  
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;  
Query Match 7.4%; Score 147.6; DB 22; Length 125401;  
Best Local Similarity 45.4%; Pred. No. 2.2e-15;  
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
```

QY 8 ACCCGGACTTCCCTCGAGCAGCACCGTCCCTACCGCGCGCGCGCGCGCGCGCGCGCTGC 67

Db 80064 ACAGGCGGACACCGACCCCGCGCTTCGGACTGACCCGCGCCCTGCTGGACGCGCAC 80123  
QY 68 TCCTCAGCTACAGGGCCCTCTTCGAGGTCTTCGTGGGACGCGTGAAGCGACTGCGT 127  
Db 80124 AACAGCGCGCCCTACGCGGACCTTCGGGCCATTCAGCCGCGCGCCCTTCGCTTCGCT 80183  
QY 128 CGAAGCTGGTCTCTCTGGGGCGCGGTGTGACCGGAGCCTTCGAGAGGGACGCGCCAG 187  
Db 80184 GGAAGGGGTCTCTGCTCGCGCGCGCGCGCCACACCGTCCGCGCGCGGATCGCCCGG 80243  
QY 188 GAACCGTGCAGCTGACCGGGATCGAGCTGAGGGCGCCAGCGATGCTGAGCGCCCTGAC 247  
Db 80244 CCGGAGAGACACCTGACCATTCGCGCTTACGAGCGCGCGCGCGACCTGCTGTCTCG 80303  
QY 248 CTGCGCGCGTCTCTGGCTGACAGCGCTCCACACCGGACAGCTCTCTTCTACCGTGGC 307  
Db 80304 TCGACTCCTGTCTCTCGCGAGGTCTCCCGCGAGCGACCGCGCGCGCGCGCGCGCTCC 80363  
QY 308 CGAAGCGGTGGACGCTTCGTGCGCGCCACCGCTCCAGAGAGGAACAGCAATGACGC 367  
Db 80364 ACCGCGACTCTCTTCCAGCTGAGTGGACCCCGCTCCAGGGCGCGCGCGCGCGCAC 80423  
QY 368 CGGCCATCTACGCGCGCGCGCTTCAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCTGATCT 427  
Db 80424 CGGCA-----CCGTCGCGCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 80474  
QY 428 TCGCGCGTGTCTGCGCGCTGACGCGCGGCGATCGCTCTCGGCGCTCGCTCGCGCTCT 487  
Db 80475 TCGCG-----CCAGCGGATCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 80525  
QY 488 ACCATGACGCGCGCGCGCTGGCTGTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547  
Db 80526 CCGAGCGGAGGGCGCGCTCCCGACCTGGTGTCAACACCGCTAC-----CACCACCGCG 80582  
QY 548 GCGCTTACAGTGTACTCTGCTGCTGCGGACGAGACGAGCGCGCGCGCGCGCGCGCGCG 607  
Db 80583 GCGCGCGCGTCTCCGACGCA 80642  
QY 608 AGCGAGACGACGCGCGCTGCGCTTCACGACACGCGCGCGCTGATGCGCGCTGCGCG 667  
Db 80643 AGTGGCTCGCGACGACCGCTTCGCGGACGCGCGCGCTGCTCTGCTACCGCGCGCGCA 80702  
QY 668 CGGACTTCTCGGTGTGCGGTGCGGCGTGTGCGGACGCGGACGAGCGGCGGCGGCTGG 727  
Db 80703 CGGAGGCGACCG 80762  
QY 728 GCGCGCGCGGTCTGCTGCGCTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787  
Db 80763 AGAACCGCGCGGTTCGCGCTTCCTGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80822  
QY 788 AGCAGTACTCTTACTGCTGCGACGCGCGGCTTCACGTTACCGCGCTGCGCGCGCGCGCG 847  
Db 80823 AGACCTGGCCACCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80879  
QY 848 TCATCGTCAACACCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 907  
Db 80880 GCACCGAGTGTGACG 80939  
QY 908 GGTTCGCGACCGCAAGATCTACTGCGGACACCGCGCGCGCGGAGAAATGGGTCTCGCTATCG 967  
Db 80940 GGAACCGCGGCG 80999  
QY 968 GAGACTCGGTGACGCGCGTTCATACCTTCGCGGAACTGCGCGACTGGACATTCGCG 1027  
Db 81000 CCGCGCACTGTGTCG 81059  
QY 1028 GCACTTGGATACCGCAACACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087  
Db 81060 CCG 81119  
QY 1088 TCACGCGAGACGCGGACGCGCGACTGCGGTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1147

Db 81120 ACATCGCGCGCTGCGAGCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81179  
QY 1148 TCGGCGTCTCCCATGAGCTAGCTACTAGCTACTGACAGGACACCTGGAGCGGAGGAGTTCACG 1207  
Db 81180 CGGCGGGGACCGCTACCGTCTGCTCCACACCGCGCGCGCTGCTGAGACGAGCGGTCC 81239  
QY 1208 CGGAGACCTACCGCGCAATGGCTGAGTGGGGTGGGACTGGTACGCGCGCGCGTCACT 1267  
Db 81240 TCGGCTCTCTACCGCT 81299  
QY 1268 GGCATCTGATCAGCG 1327  
Db 81300 GGCATCTGACGAGCG 81359  
QY 1328 AGTAGCG 1387  
Db 81360 TCG 81419  
QY 1388 TCGTCTCGGAGTGGGGCTGCG-----CGGACGCGCTGGGCGCTGTACACCTCTCTGAGCAC 1445  
Db 81420 AGCGCTCTCG 81479  
QY 1446 CCGCTGGGCGCGCTGACGAACTAGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1505  
Db 81480 CGTGACCG 81537  
QY 1506 CGAGCG 1565  
Db 81538 CGCGCGCTCG 81597  
QY 1566 CTGGGACACCG 1622  
Db 81598 CTTGGCG 81657  
QY 1623 GCACGAGACATCGGCAAGTAGCGGACGACCTGTACGTGCGCGGAGGACCTTCCGACCT 1682  
Db 81658 CGAGGCGACATCG 81717  
QY 1683 CGCGCGGTACTCGCTCG 1736  
Db 81718 CGCGCGGTCTCGGAGCG 81777  
QY 1737 CGCGCGGTCTCGGAGCG 1796  
Db 81778 CGCGCGCGCGGAGCGCGCTCTGGAATCTGTCGCGACCGCGCGCGCGCGCGCGCGCGCG 81837  
QY 1797 CG 1856  
Db 81838 CG 1897  
QY 1857 GCTTACACCG 1904  
Db 81898 CACCGCGTGAATCG 81945

## RESULT 9

AA53491/C

ID AAX53491 standard; DNA; 114955 BP.

XX AAX53491;

AC AAX53491;

XX AAX53491;

DT 05-JUL-1999 (first entry)

XX Human adenosine A1 receptor antisense oligonucleotide fragment.

XX Antisense oligonucleotide; multiple target; antisense treatment;

KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW acute asthma; allergy; asthma; impeded respiration;

KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KW colon cancer; breast cancer; lung cancer; pancreatic cancer;



KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
KW prostate cancer; ss.

OS Synthetic.

XX WO9913886-A1.

PN

XX

PD

XX

PF

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

The specification describes antisense oligonucleotides (AA52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences AA5272-74. These multiple target oligonucleotides (specifically AA55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, CC hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.

XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 6.7%; Score 134.8; DB 20; Length 114955;

Best Local Similarity 31.6%; Pred. No. 2.4e-13;

Matches 617; Conservative 180; Mismatches 1134; Indels 20; Gaps 5;

QY 32 CCGTCCCTACCGGCGACCGCCGCGGCGGCTGCTCTCAGCTAGACGGCCCTCC 91

DB 105241 CCGGCGCGCGCCNNHNNNSGCGCGCGCGCGCGCGCCNNHNNNSGCGCGCG 105182

QY 92 TCGAGTCTTCGTCGGGACGCTGAGGCGACTCGCTCGAACCTGPTCTCTGGGGCG 151

DB 105181 CGCGCGCGCCVNNHNNNSGCGCGCGC--CGGCGCGCGCCVGNHNNNSGCGCGCG 105124

QY 152 CGGCTGTACCGGACGCTCGAGACGGACGGCCAG-GAACTGTGACGTGACCGGATC 210

DB 105123 CCGGCGCGCGCCVGNHNNNSGCGCGCGCGCGCGCGCCVGNHNNNSGCG 105064

QY 211 GACGTGAGCGCGCCAGCATGTGACGCCCTGAACTTCCGCGCGCTTCTGGGCTGACCA 270

DB 105063 CCGGCGCGCGCGCCVGNHNNNSGCGCGCGCGCGCGCGCCVGNHNNNSGCG 105004

QY 271 GCGCTCCACCGACAGCTCTCTCTTATCCGCTGCGGACACAGGCTGAGCTTCGTC 330

DB 105003 NNHNNNSGCGCGCGCGCGCGCGCCVGNHNNNSGCGCGCGCGCGCGCGCG 104944

QY 331 GCGCCACCGCTCCACGAGAGGACCAATGACCGCGCGCATCTCAGCGCGCGCGCTG 390

DB 104943 GCCVCGCGCGVGNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104884

QY 391 CTCCAGGAGCGCGCGGAGACACTCGCCCTGATCTTCTGGCGGTGCTGTGCGCGCTGCA 450

DB 104883 CVGCGNNHNNNSVGGCCVCGCGNNHNNNSVGGCCVCGCGNNHNNNSVGGCCVGGCG 104824

QY 451 GCCCGGCGATCCGCTCCGGGCTCGCTCCGCTGCGCTTACCATGACGCGCGCGCGCGCG 510

DB 104823 NNHNNNSCGCGCGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104764

QY 511 TGCGCTCGGACCG 570

DB 104763 CGNNHNNNSGCG 104704

QY 571 CACTCCGACGACGACG 630

DB 104703 CG 104644

QY 631 GCGTTCACGACGACG 689

DB 104643 GCGVCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104584

QY 690 GTCGCGCGCTGTCG 749

DB 104583 VCGCGNNHNNNSGCG 104524

QY 750 GCGGACCG 809

DB 104523 CVGCGNNHNNNSGCG 104464

QY 810 CGAGCGCGGTTTACGTTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 869

DB 104463 CVGCGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104404

QY 870 CG 929

DB 104403 CG 104344

QY 930 CTGGGACACCGCGCGCGGAGATGGTCTCGCTCATCGGACGACTCGGCTACGCCCGT 989

DB 104343 NNSGCG 104284

QY 990 C-----TACACTCGCGCGAACTCGCGGACTGAGACACTTTCGCGCGCACTTCGAC 1038

DB 104283 CVGCGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104224

QY 1039 TACCGAAGCG 1098

DB 104223 CGCGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104164

QY 1099 GACGGGA-----CACGCCACTGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153

DB 104163 GACGNNHNNNSGCG 104104

QY 1154 TCCCATGACGTACGCTTACTGAGACGACGACCTGGGACGCGGAGAGTTCACGCGCGAG 1213

DB 104103 CCGGCGCGCGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104044

QY 1214 ACCTCACCGCGCAATGGCTCGACTGGGCTGGGACTGGTACGCGCGCGCTCACTGGCCAT 1273

DB 104043 NNHNNNSGCG 103984

QY 1274 CGATGACG 1333

DB 103983 CCGGCG 103924

QY 1334 CGGACGCGAGCTTCCCG 1393

DB 103923 GCG 103864

QY 1394 GCGAGCTGCGGCTCG 1453









Db 21093 CTGCGGCGGCGGCGGCGGCTACGAGAGTCTCTGCGCACCATCCGCGCGCGCG 21152  
QY 1689 GTAATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1748  
Db 21153 CGAGGAACCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21212  
QY 1749 GCACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1808  
Db 21213 GCGGCTGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21272  
QY 1809 CGTGTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1868  
Db 21273 CGGCTACAGATCCCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21332  
QY 1869 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1880  
Db 21333 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21344

## RESULT 12

AAD17185  
ID AAD17185 standard; DNA; 27541 BP.

AC AAD17185;

DT 29-NOV-2001 (first entry)

DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; nys2; ds.

OS Streptomyces noursei.

FH Key Location/Qualifiers  
FT CDS complement (454..1191)

FT /\*tag- a

FT /product- "NysF protein"

FT /note- "CDS does not include start codon"

FT complement (1275..3092)

FT /\*tag- b

FT /product- "NysG protein"

FT complement (3070..4824)

FT /\*tag- c

FT /note- "CDS does not include start codon"

FT /product- "NysH protein"

FT 5122..6156

FT /\*tag- d

FT /product- "NysD3 protein"

FT 6338..27541

FT /\*tag- e

FT /product- "NysI partial protein"

FT /note- "CDS does not include stop codon"

PN WO200159126-A2.

PN 16-AUG-2001.

PD 08-FEB-2001; 2001WO-GB00509.

PF 08-FEB-2000; 2000GB-0002840.

PR 10-APR-2000; 2000GB-0008786.

PR 14-APR-2000; 2000GB-0009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

PA (SNF) SINTEF STIFTELSEN IND TEK FORSK.

PA (ALPH-) ALPHARMA AS.

PA (SINV-) SINVENT AS.

PA (DZIE-) DZIEGLEWSKA H.

PA (ZOTC-) ZOTCHEV S B.

PA (SEKU-) SEKUROVA O N.

PA (FJAE-) FJAERVIK E.

PA (BRAU-) BRAUTASET T.

(STRO/) STROM A R.

XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI: 2001-557614/62.

DR P-PSDB: AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.

XX New nystatin polyketide synthase polynucleotides and polypeptides,

PT useful as antibiotics and antifungals -

XX Claim 2: Page 151-166; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present

CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

QY Query Match 5.5%; Score 109.2; DB 22; Length 27541;

Db Best Local Similarity 43.5%; Pred. No. 3.2e-09;

QY Matches 859; Conservative 0; Mismatches 1088; Indels 27; Gaps 7;

Db 4 GTGACCCCGGACTTCCCTTCGAGACACCGTCCCTTACCGGCGGCGGCGGCGGCGG 63

Db 8285 GTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8344

QY 64 CTGCTCTCAGCTAGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123

Db 8345 GGGGTCTCTCCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8404

QY 124 GCGTCGAACCTGCTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 183

Db 8405 GCGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8464

QY 184 CCAGGAACCGTGCAGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 243

Db 8465 CTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8524

QY 244 GAACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303

Db 8525 GTCTGTCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8584

QY 304 TGCCCGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 363

Db 8585 CGCAAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8644

QY 364 AGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 423

Db 8645 CTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8704

QY 424 ATCTTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 483

Db 8705 GTCTTCAACCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8764

QY 484 GTCTACACATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 543

Db 8765 GTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8824

QY 544 CACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

Db 8825 GACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8884

QY 601 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660

Db 8885 TGCTTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8944

QY 661 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714

Db 8945 GAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9004

[illegible]



KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
KW prostate cancer; ss.  
OS Synthetic.  
XX WO9913886-A1.  
PN 25-MAR-1999.  
XX 17-SEP-1998; 98WO-US19419.  
XX 09-JUN-1998; 98US-0093972.  
PR 17-SEP-1997; 97US-0059160.  
XX (UYEC-) UNIV EAST CAROLINA.  
PA Nyce JW;  
PI WPI; 1999-229400/19.  
XX New antisense oligonucleotides used in treatment of, e.g. pulmonary  
PT vasoconstriction  
XX Disclosure; Page 37; 120pp; English.  
XX The specification describes antisense oligonucleotides (AA552869-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene  
CC initiation codons, genomic flanking regions, intron-exon borders, the  
CC 5'-end, the 3'-end and the juxta-section between coding and non-coding  
CC regions and all segments of RNAs encoding proteins associated with one  
CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
CC may be derived from sequences AA55272-74. These multiple target  
CC oligonucleotides (specifically AA55180-271) can be used for the  
CC antisense treatment of diseases and conditions. Typical diseases and  
CC conditions are those associated with impaired respiration and  
CC inflammation, including lung diseases, pulmonary vasoconstriction,  
CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impaired  
CC respiration, respiratory distress syndrome, pain, cystic fibrosis,  
CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic  
CC obstructive pulmonary disease (COPD), and cancers such as leukemias,  
CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,  
CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,  
CC hepatic metastases, as well as all types of cancers which may metastasize  
CC or have metastasized to the lungs, including breast and prostate cancer.  
XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;  
SQ  
Query Match 5.0%; Score 99.8; DB 20; Length 114955;  
Best Local Similarity 34.0%; Pred. No. 8.9e-08;  
Matches 604; Conservative 118; Mismatches 1032; Indels 21; Gaps 7;  
QY 2 CGGTGACACCCCGACTTCCCTCGACGACACACCTCCCTTACCGCGCGACCGCGCGCC 61  
DB 104356 CBGGGCGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGGGSNNN 104415  
QY 62 GACTGTCTCTACGCTAGACGGCGCCCTCTCTCGAGGTCTTCTGTCGGGAGCGGTGAGCGGA 121  
DB 104416 DNNCCGCBGGCGCGCGCGCGGCGSNNNDNCCBGGCBGGCGCGCGCGCGCG 104475  
QY 122 CTCGTGCAACCTGTCTCTCTGGGCGCGCGCGGTGTACCGGAGCCTGTGACGACGCGAC 181  
DB 104476 GCCSNNNDNCCBGGCGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCG 104535  
QY 182 GGCAGGAAACCGTGCAGTACCGCGATCGACGTGAGGCGCGCGCGCGATGCTGACGCGCC 241  
DB 104536 CGCGGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGG 104595  
QY 242 CTGAACCTCGCGCTTCTGGGTGACGAGCGCTCCACCGCGCGCGCGCGCGCGCGCGTCTTCTTACC 301  
DB 104596 CGCGGGSNNNDNCCBGGCGCGCGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGCGCG 104655  
QY 302 GCTGCCGAACAGGTGTGACGCTTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361

DB 104656 SNNNDNCCBGGCBGGCGCGCGGGSNNNDNCCBGGCBGGCGCGCGGGSNNNDNCC 104715  
QY 362 TGAGCGCGCGCATCTCAGCGCGCGCGCGTGTCTCCAGGAGCGCGCGCGCGCGCGCG 421  
DB 104716 CGBGGCBGGCGCGGSNNNDNCCBGGCGCGCGGCGGSNNNDNCCBGGCGCGCGG 104775  
QY 422 TGATCTTTCGGCGGTGTCTGTCGCGCGCTCAGCCCGCGGCATCGCTCCGGGTTCGCTCGGTG 481  
DB 104776 SNNNDNCCBGGCBGGCGGSNNNDNCCBGGCBGGCGGSNNNDNCCBGGCGCGCGG 104835  
QY 482 CCGTCTACCATGACCGCGCGCGCGCGTGTCTTCGACCGCGCGCGCGCGCGCGCGTACCA 541  
DB 104836 SNNNDNCCBGGCBGGCGGSNNNDNCCBGGCGCGGCGGSNNNDNCCBGGCGCGG 104895  
QY 542 CCCACGCGCGCTACCACTGTACTCTCAGCTCCGACCGACGAGCAACAGCGCGCGCGGT 601  
DB 104896 CBGGCBGGCG 104955  
QY 602 GGGACACGCGAGCAGCAGCGCGCGCGTCTCAGCGACCGCGCGCGCGCGCGCGCGG 658  
DB 104956 GGGCGGSNNNDNCCBGGCBGGCGCGCGCGCGCGCGCGCGGSNNNDNCCBGGCGCGG 105015  
QY 659 CGCTGCGCGCGCGCTTCCCGTGTGTCTCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 718  
DB 105016 CG 105075  
QY 719 CAGGTTTCGGCGCGCGCGCGCGTGTCTGCGTCTCGGACCGCGCGCGCGCGCGCGCG 778  
DB 105076 CBGGCGCGCGCGCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCGCGCG 105133  
QY 779 AGTACGAGGAGTACTCTTACTGTCTGACGCGCGCGGTTCACGTTTCCCGCGCGCGTGC 838  
DB 105134 NNDNCCBGGCGCGCGCGCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCG 105193  
QY 839 CCACCGCGTCTGTCTAACACCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 898  
DB 105194 NNDNCCBGGCGCGCGCGCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCG 105253  
QY 899 ACCCGAGTGTTCGGCG 958  
DB 105254 DNNCGCGCGCGCGCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCGCGCG 105312  
QY 959 CGCTATCGGACGACTGCGGTACGCGCGGTCTTCTACACTCTCGCGGAACTCTCGGCG 1018  
DB 105313 CGCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCGCGCGCGCGCGCGCG 105372  
QY 1019 CACTTCCCGCACTTCTGACTTACCGCAACCGCGCGTGTGTGTGTGTGTGTGTGTGTGT 1078  
DB 105373 CGSNNNDNCCBGGCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCGCGCG 105432  
QY 1079 TGTTCGAGATCACCGCAGACGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1136  
DB 105433 GCGCGSNNNDNCCBGGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCGCGCG 105492  
QY 1137 CGCTACGCGATCGGCTTCCCATGACGTACGCTTACTTGACAGGACGCGCGCGCGCG 1196  
DB 105493 NNDNCCBGGCGCGGTCTCCGGSNNNDNCCBGGCGCGCGCGCGCGCGCGCGCG 105552  
QY 1197 GCAGTTTCACCG 1256  
DB 105553 GCGCGCGCGCGCGGSNNNDNCCBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105612  
QY 1257 GCGCGTCACTGCGCATCGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1316  
DB 105613 CGCG-----CGSNNNDNCCBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105667  
QY 1317 GAACAACTGGAAGTACG 1376  
DB 105668 CCSSNNNDNCCBGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 105727  
QY 1377 GCAGAACTGATGTCGCGGAGCTGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1431



```
Db 105728 GGGCCCCCTGGCTCGGCCCGCCCGGGCCCGGCTTGGCCGCGCCGCSNNNDNNGGCCCCCTGGC 105787
Qy 1432 ACCCTCTCTGAGCACCCCGTGGCAGCGCTGACGAACCTAGCTACCGCCGACACACACACTC 1491
Db 105788 TCGGCCCGCCGCGCGGCTTGGCCCGCCGCGGNNNDNNGGCGCCCTGGCTCGGCCCGCGG 105847
Qy 1492 CCCGACCGGACCGTGCAGCGGAGCGCGCTCTCTGCCATGG---AACGGACGCGCATACGAG 1548
Db 105848 CCGCGCTTGGCCCGCCGCSNNNDNNGGCGCCCTGGCTCGGCCCGCCGCGGCTTGGCCCG 105907
Qy 1549 ATCGAGCTCGACATCGCTGGGACACCGGAGACGCTGCGGCATCTCGTGGCGGCGCTCC 1608
Db 105908 CCCNNNDNNGGCGCCCTGGCTCGGCCCGCGCGGCTTGGCCGCSNNNDNNGGCGC 105967
Qy 1609 CCGACGGAACCCGCGCACACGACGCAAGTACGAGCAGACCTGCTACGTCGACCGA 1668
Db 105968 CTTGGCTCGGCCCGCGCGCGCTTGGCCGCSNNNDNNGGCGCCCTGGCTCGGCCCGCGC 106027
Qy 1669 GGACCCCTCGGACCTCGCGGCTACTCGCTCGCCCGCTACTCGGAGCGCGCCCGCCCATC 1728
Db 106028 GGCCCGGCTTGGCCGCSNNNDNNGGCGCCCTGGCTCGGCCCGCGCGGCTTGGCCSNN 106087
Qy 1729 GACCCCGCGCGCCGATCGCTGACCTGCGCATCCT 1763
Db 106088 NDNNGGCGCCCTGGCTCGGCCCGCGCGGCGCGGCTT 106122

RESULT 15
AAAI4651
ID AAI4651 standard; DNA; 77536 BP.
XX AC AAI4651;
XX DT 08-AUG-2000 (first entry)
XX DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW FK-506 binding protein; polyketide compound; transplant rejection;
KW graft-versus-host disease; uveitis; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX OS Streptomyces hygroscopicus.
XX FH Key Location/Qualifiers
FT CDS complement (412..1836)
FT FT /*tag= a
FT FT /note= "fkBW gene"
FT CDS complement (2020..3579)
FT FT /*tag= b
FT FT /note= "fkBV gene"
FT FT 3969..4496
FT FT /*tag= c
FT FT /note= "fkBR2 gene"
FT CDS complement (4595..5488)
FT FT /*tag= d
FT FT /note= "fkBR1 gene"
FT FT 5601..6818
FT FT /*tag= e
FT FT /note= "fkBE gene"
FT CDS 6808..8052
FT FT /*tag= f
FT FT /note= "fkBF gene"
FT CDS 8156..8824
FT FT /*tag= g
FT FT /note= "fkBG gene"
FT CDS complement (9122..9883)
FT FT /*tag= h

/note= "fkBH gene"
complement (9894..10994)
/*tag= i
/note= "fkBI gene"
complement (10987..11247)
/*tag= j
/note= "fkBJ gene"
complement (11244..12092)
/*tag= k
/note= "fkBK gene"
complement (12113..13150)
/*tag= l
/note= "fkBL gene"
complement (13212..23988)
/*tag= m
/note= "fkBC gene"
complement (13452..13662)
/*tag= n
/note= "ACP6"
complement (13761..14394)
/*tag= o
/note= "KR6"
complement (14517..15294)
/*tag= p
/note= "ER6"
complement (15438..16587)
/*tag= q
/note= "dehydratase domain (DH) 6"
complement (16587..17820)
/*tag= r
/note= "acyltransferase domain (AT) 6"
complement (17820..19053)
/*tag= s
/note= "KS6"
complement (19116..19326)
/*tag= t
/note= "ACP5"
complement (19464..20097)
/*tag= u
/note= "KR5"
complement (20241..21420)
/*tag= v
/note= "DH5"
complement (21420..22653)
/*tag= w
/note= "AT5"
complement (22653..23892)
/*tag= x
/note= "KS5"
complement (23992..46573)
/*tag= y
/note= "fkBB gene"
complement (24163..24373)
/*tag= z
/note= "ACP4"
complement (24997..26146)
/*tag= aa
/note= "DH4 (inactive)"
complement (26146..27430)
/*tag= ab
/note= "AT4"
complement (27430..28684)
/*tag= ac
/note= "KS4"
complement (28750..28960)
/*tag= ad
/note= "ACP3"
complement (29092..29740)
/*tag= ae
/note= "KR3"
complement (29869..31018)
/*tag= af
/note= "DH3 (inactive)"
```





```

Db 62147 GGCCTCGGCTCCGGGAGGACTTCGTGATGGCCCGCGGATGGACCCCGGCACAGCCGAT 62206
Qy 1371 CAACGGGAGAACTGG---ATCGTCCGGAGCTGCGGCTCGCCCGACAGCCCTGGCGGCTG 1427
Db 62207 GACCGGCTCCGTACCGCCCATCTGAGCGGCTGCGAGGAGCGCGCGCTGCGCCG 62266
Qy 1428 GTACACCTCTCTGAGCACCCCGTGGCAGCGCTGACGAACCTACGTACCGGCCACCAAC 1487
Db 62267 TGCCGGGAGAGCTTCGCCCAAGCGCTCGCGGAGCTGCCCGACGCGGACCGCGGCGGC 62326
Qy 1488 ACTCCCGGACCGGACCGCTCGAGCGGAGCGGCTGCTGCGCATGGACGCGGATACGA 1547
Db 62327 GCTGACCACTCTGCTCGGAGCGCCAGCGCGCGCTGCTCGGCCACGCGGACGCGCTCGA 62386
Qy 1548 GATCGAGCTCGACATCGCTCGGACACCGGACGCAACGTCGGCATCTCGGTGGGCGGCTC 1607
Db 62387 GATCGCGCGGACCGACGACGTTCAAGGACCTCGGCATCGACTCGCTCACCGCGATCGACT 62446
Qy 1608 CCCCAGCGAACCAGCGGACACGACATCGGGAAGTACGAGGAGACC---TGTAGTCTGA 1664
Db 62447 GCGCAACCGGCTCGCGGAGGCGGACCGGGCTGCGGCTGAGTGCCACGCTGGTGTTCGACCA 62506
Qy 1665 CCGAGGACCTCCGACCTCGCGGGTACTCGCTCGCCCTACTCGCGAGCGCGCGCC 1724
Db 62507 CCGGACCTCGGGTCTCGCGCGCCCAAGCTCGGACCGATCTGTTCGGACGCGCGGTGCC 62566
Qy 1725 CATCGACCCCGGCGCCGATCGGTGACCTGCGCATCTCGTCGACACCCAGAGCGTCTGA 1784
Db 62567 CACGCCCGGCGGAGCGGACGACGCCACCGACGAGGCACTCGCGATCGTGGCATGGC 62626
Qy 1785 GGTCTTCGTCAACCGCGCGGCGGACCGGTCTTCCAGAGGTCTACTTCGCGCGAGGGCGA 1844
Db 62627 GTGCGGACTGCGCGGCGGGTGGCTCGCCCGGAGGACCTGTGGCAGCTCGTGGCGTCCGG 62686
Qy 1845 CACGG 1849
Db 62687 CACCG 62691

```

Search completed: May 10, 2003, 06:29:26  
 Job time : 4644.5 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 04:44:16 ; Search time 2890 Seconds  
(without alignments)  
11213.553 Million cell updates/sec

Title: US-09-868-328B-2

Perfect score: 2001

Sequence: 1 gcggtgaccccgacttccc.....atcgacacgtctctcgctt 2001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_man:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Length | DB ID   | Description |
|------------|-------|--------------|---------|-------------|
| 1          | 94.4  | 4.7          | 1288 14 | BQ678719    |
| C 2        | 90.6  | 4.5          | 1016 17 | CNS03L09    |
| C 3        | 88.2  | 4.4          | 1046 14 | BQ952554    |
| C 4        | 88    | 4.4          | 1651 12 | BG809816    |
| 5          | 86.6  | 4.3          | 1065 14 | BQ681076    |
| 6          | 86.6  | 4.3          | 1218 14 | BQ876717    |

|      |      |     |      |    |          |
|------|------|-----|------|----|----------|
| 7    | 84.4 | 4.2 | 1339 | 13 | BM458211 |
| C 8  | 84.2 | 4.2 | 827  | 17 | AZ183795 |
| 9    | 82.6 | 4.1 | 733  | 17 | AG039579 |
| C 10 | 82   | 4.1 | 934  | 17 | AG046262 |
| C 11 | 82   | 4.1 | 1023 | 17 | AG128304 |
| C 12 | 81.6 | 4.1 | 1341 | 17 | AG030611 |
| 13   | 81.2 | 4.1 | 1242 | 14 | BM911414 |
| 14   | 81.2 | 4.1 | 1281 | 12 | BG852363 |
| 15   | 80.8 | 4.0 | 1051 | 14 | BQ689718 |
| C 16 | 80.6 | 4.0 | 1116 | 12 | BG810038 |
| C 17 | 80.6 | 4.0 | 1233 | 17 | AG074706 |
| C 18 | 80.4 | 4.0 | 1321 | 17 | AG126084 |
| 19   | 79.8 | 4.0 | 1798 | 17 | AG171124 |
| C 20 | 79.4 | 4.0 | 1112 | 14 | BQ058195 |
| C 21 | 78.2 | 3.9 | 666  | 12 | BG786337 |
| 22   | 78.2 | 3.9 | 724  | 17 | AG171201 |
| 23   | 78.2 | 3.9 | 1061 | 17 | AG046200 |
| 24   | 78.2 | 3.9 | 1150 | 14 | BQ892481 |
| 25   | 78   | 3.9 | 775  | 17 | AZ185516 |
| 26   | 78   | 3.9 | 821  | 17 | AG063084 |
| C 27 | 78   | 3.9 | 1243 | 14 | BM913931 |
| C 28 | 77.8 | 3.9 | 1143 | 10 | AW731158 |
| C 29 | 77.6 | 3.9 | 715  | 17 | AG036665 |
| 30   | 77.6 | 3.9 | 1128 | 14 | BQ710735 |
| C 31 | 77.6 | 3.9 | 1170 | 17 | AG111669 |
| C 32 | 77.2 | 3.9 | 1297 | 14 | BQ064843 |
| C 33 | 77   | 3.8 | 1025 | 17 | AG137560 |
| C 34 | 77   | 3.8 | 1123 | 17 | AG080476 |
| C 35 | 77   | 3.8 | 1134 | 17 | AG043044 |
| C 36 | 77   | 3.8 | 1137 | 12 | BG809979 |
| 37   | 76.8 | 3.8 | 1268 | 13 | BM551006 |
| C 38 | 76.6 | 3.8 | 1127 | 14 | BQ718629 |
| 39   | 76.4 | 3.8 | 763  | 17 | AG083422 |
| C 40 | 76.4 | 3.8 | 1197 | 13 | BI416470 |
| C 41 | 76.2 | 3.8 | 1138 | 9  | AI374006 |
| 42   | 76   | 3.8 | 1149 | 14 | BM910848 |
| C 43 | 76   | 3.8 | 1798 | 17 | AG171124 |
| C 44 | 75.8 | 3.8 | 1250 | 13 | BM560215 |
| C 45 | 75.8 | 3.8 | 1424 | 14 | BQ720687 |

ALIGNMENTS

|            |  |             |              |            |                 |
|------------|--|-------------|--------------|------------|-----------------|
| RESULT 1   | BQ678719   | 1288 bp     | mrna         | linear     | EST 15-JUL-2002 |
| LOCUS      | AGENCOURT_8199257  | NIH_MGC_112 | Homo sapiens | cdna clone | IMAGE:6261563   |
| DEFINITION | 5', mRNA sequence.   |             |              |            |                 |
| ACCESSION  | BQ678719   |             |              |            |                 |
| VERSION    | BQ678719.1   |             |              |            |                 |
| KEYWORDS   | EST.   |             |              |            |                 |
| SOURCE     | human.   |             |              |            |                 |
| ORGANISM   | Homo sapiens   |             |              |            |                 |
| REFERENCE  | 1 (bases 1 to 1288)  |             |              |            |                 |
| AUTHORS    | NTH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .  |             |              |            |                 |
| TITLE      | National Institutes of Health, Mammalian Gene Collection (MGC)   |             |              |            |                 |
| JOURNAL    | Unpublished (1999)   |             |              |            |                 |
| COMMENT    | Contact: Robert Strausberg, Ph.D.<br>Email: <a href="mailto:cgabs-r@mail.nih.gov">cgabs-r@mail.nih.gov</a><br>Tissue Procurement: DCTD/DTF<br>CDNA Library Preparation: Rubin Laboratory<br>CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)<br>Clone distribution: Agencourt Bioscience Corporation<br>found through the I.M.A.G.E. Consortium/LLNL at:<br><a href="http://image.llnl.gov">http://image.llnl.gov</a><br>Plate: LLCM2422 row: h column: 12<br>High quality sequence stop: 171. |             |              |            |                 |
| FEATURES   | Location/Qualifiers  |             |              |            |                 |
| source     | 1. 1288  |             |              |            |                 |









/clone\_lib="NIH\_MGC\_110"  
/tissue\_type="ductal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
Site="Organ: pancreas; Vector: pOMB7; Site.1: XhoI;  
Site.2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
BASE COUNT 63 a 629 c 334 g 95 t 97 others  
ORIGIN

Query Match 4.3%; Score 86.6; DB 14; Length 1218;  
Best Local Similarity 41.9%; Pred. No. 3.1e-05;  
Matches 334; Conservative 0; Mismatches 463; Indels 1; Gaps 1;  
QY 7 CACCCGGACTTCCCTCGACGACACCGTCCCTACCGCGCGACGCGCGCGCGACTG 66  
DB 419 CCCCCGGGGCGCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478  
QY 67 CTCCTCAGCTAGACGGCGCTCTCTCGAGGTCTTCGTGGGAGCGGTGAGCGACTCG 126  
DB 479 CCCCCNCGCNCNCCCG 538  
QY 127 TCGAACCTGTGCTCTCTGGGGCGCGCGGTGTGACCGCGAGCTCGAGACGGCGCGCA 186  
DB 539 CCCCCCG 598  
QY 187 GGAACCGTGTGACGCGGATGACGTGTGAGGGCGCGCGCGCGCGCGCGCGCGCGCG 246  
DB 599 CCG 658  
QY 247 CCGCGCGCGGTCTGGGTGTGAGCGGTCTCCACCGCGACACTCTCTTACGCGTGC 306  
DB 659 CGCGCCNCCCG 718  
QY 307 CGGAACAGGGTGTGAGCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366  
DB 719 CG 778  
QY 367 CGGGCATCTACGCGCGCGCGGTGTCTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 426  
DB 779 CCCCCCG 838  
QY 427 TTCGGCGGTGTGTGCGCGGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486  
DB 839 CGCNCGCGNCCCG 898  
QY 487 TACCAATGACCG 546  
DB 899 CCCCCCG 958  
QY 547 GCGCGCTACCACTGTACTGTACTGTCTCGACAGACACGCGCGCGCGCGCGCGCGCG 606  
DB 959 GCG 1018  
QY 607 CACGGAGCACACGCGCGGTGCGCTTTCAGCGACACGCGCGCGCGCGCGCGCGCGCG 666  
DB 1019 CGCGCCCCCG 1077  
QY 667 CCGCGACTTCCCGGTGTGGGTTCGGGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726  
DB 1078 CG 1137  
QY 727 GCGCGCGCGCGGTGTGCGCGCTCGGACCGACGCGCGCGCGCGCGCGCGCGCGCG 786  
DB 1138 GCG 1197  
QY 787 GAGCAGTACTTCTACTGG 804  
DB 1198 GCGCGCGCGCGCGCGCGCG 1215

## RESULT 7

BM458211

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM458211 1339 bp mRNA linear EST 05-FEB-2002  
AGENCOURT\_6411296 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5530584  
5', mRNA sequence.  
BM458211  
BM458211.1 GI:18507251  
EST.  
human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1339)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM12210 row: o column: 01  
High quality sequence stop: 134.  
Location/Qualifiers

1..1339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5530584"  
/clone\_lib="NIH\_MGC\_71"  
/tissue\_type="leiomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="organ: uterus; Vector: pCMV-SPORT6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 2.1 kb."  
63 a 780 c 403 g 78 t 15 others

## BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

0; Mismatches

426; Indels

24; Gaps

5;

QY

2

CGGTGACCGCGACTTCCTCTGACGACACCGTCCCTACCGCGCGCGCGCGCGCGCGCGCG 61

DB

253

CGGGCG 312

QY

62

GACTGCTCTCAGCCTAGACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121

DB

313

GCG 372

QY

122

CTGCGTGAACCTGTGCTCTGCGGGCGCGCGGTGTGACCGCGCGCGCGCGCGCGCGCG 181

DB

373

CG 432

QY

182

GCGCAGGACCGTGTGACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 241

DB

433

GCG 492

QY

242

CTGAACCTGCGCGCGGTCTGCGGTGACGAGCGCTCCACCGCGCGCGCGCGCGCGCGCG 301

DB

493

CG 544

QY

302

GCTGCCCGAACCAGGGTGTGACGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361

DB

545

GCG 604

QY

362

TGACG 421





REFERENCE  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 733)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKSI45  
R.Site 1 : SacI  
R.Site 2 : SacI  
FEATURES  
source Location/Qualifiers  
1..733  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-016L16.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 30 a 363 c 310 g 17 t 13 others  
ORIGIN  
Query Match 4.1%; Score 82.6; DB 17; Length 733;  
Best Local Similarity 48.3%; Pred No. 0.00014;  
Matches 287; Conservative 0; Mismatches 299; Indels 8; Gaps 2;  
QY 157 GTGACCGGAGCTCGAGCGCAGCGGCGAGGACCGTGCAGTGCAGCGATCGACGTC 216  
Db 140 CGCACCAGCG 199  
QY 217 GAGCGCGCAGGATGCTGACGCGCTGAACCTGCGCGCGTTCGGGTGACGAGCGTC 276  
Db 200 GNGGGGGTGGGGAAGAACCGCGCGCTCCCTTGGCGCGCGCGCGCGCGCGCGCG 259  
QY 277 CCACCGCCGACGCTCTCTTACGCGTCG---CCGAACAGGCTGACGCTTC-----G 328  
Db 260 CG 319  
QY 329 TCGCGCGCCACCGCTCCACGAGAGGAAACAGCAATGACGCGCGCCATCTCACCGCGCG 388  
Db 320 CG 379  
QY 389 TGCTCCAGGAGCG 448  
Db 380 CG 439  
QY 449 CAGCGCGCGCATCGCTCGCGCGCTCGCTCGCTGCTTACCATGACGCGCGCGCGCG 508  
Db 440 CG 499  
QY 509 GCTGGCTTGGACCGCCCAACCGCGCGTTCACACCGCGCGCGCGCGCGCGCGCGCG 568  
Db 500 CGCGAGCGGGGACCG 559  
QY 569 TCGACTCGACAGAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628  
Db 560 CG 619  
QY 629 TCGCGCTTACGACGACGCGCGGATGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 688  
Db 620 CCG 679

QY 689 GGTGCGCGCGTCTGTCGCAACCGCAACACGCGAGGTTGCGCGCGCGCGCGTGC 742  
Db 680 GGGCG 733  
RESULT 10  
AG046262/c  
LOCUS Pan troglodytes DNA, clone: PTB-025G21.R, genomic survey sequence.  
DEFINITION AG046262  
ACCESSION AG046262  
VERSION AG046262.1 GI:15583154  
KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male  
BAC Library clone:PTB-025G21.R.

ORGANISM  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
REFERENCE  
AUTHORS 2 (bases 1 to 934)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.

COMMENT  
PRIMERS  
Sequencing: M13Rev  
LIBRARY  
Vector : pKSI45  
R.Site 1 : SacI  
R.Site 2 : SacI  
FEATURES  
source Location/Qualifiers  
1..934  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-025G21.R"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 30 a 201 c 654 g 16 t 33 others  
ORIGIN

Query Match 4.1%; Score 82; DB 17; Length 934;  
Best Local Similarity 44.8%; Pred No. 0.00018;  
Matches 301; Conservative 0; Mismatches 369; Indels 2; Gaps 1;

QY 276 CCCACCGCGCAGCTCTCTTACGCTGCCCGCGCGCGCGCGCGCGCGCGCGCG 335  
Db 917 CCCCCCG 858  
QY 336 CACCGCTCCACGAGAGAACAGCAATGACGCGCGCGCGCGCGCGCGCGCGCG 395  
Db 857 NCCTCCCG 798  
QY 396 GGGAGCG 455  
Db 797 CG 738  
QY 456 GGCATCCGCTCGCGCTCGCTCGCTTACCATGACGCGCGCGCGCGCGCGCGCG 515  
Db 737 CCGNGCG 678  
QY 516 CTGCGACCG 575



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
PRIMERS  
LIBRARY  
FEATURES  
source

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of Library PTB  
Unpublished  
2 (bases 1 to 1341)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimbes@sc.riken.go.jp, URL: http://hdp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

Sequencing: -21M13  
Vector : pKS145  
R Site 1 : SacI  
R Site 2 : SacI  
Location/Qualifiers  
1. 1341  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-003A20.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 48 a 476 c 700 g 14 t 103 others  
ORIGIN

Query Match 4.1%; Score 81.6; DB 17; Length 1341;  
Best Local Similarity 44.8%; Pred. No. 0.00022;  
Matches 388; Conservative 0; Mismatches 467; Indels 11; Gaps 5;

QY 36 CCCCCTACCGGCGGACCGCCGCGCGGCTGCTCCCTACGCTAGAGCGGCGCCCTCTCGA 95  
DB 1299 CGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1240  
QY 96 GTGCTTCGTCGGGGACGGTGAGCGACTCGGTGGAACCTGGTCTCTCTGGGGCGGCGG 155  
DB 1239 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1180  
QY 156 TGTGACCGGAGCCTGAGAC---GGCAGCGCAGGACCGTGCACGTGACCGGATCG 211  
DB 1179 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1120  
QY 212 ACGTCGAGCGCGGCGGCGGATGCTGACGCGCCCTGAACCTGCCGCGGTTCTGGGCTGACGAG 271  
DB 1119 GCGGCG 1060  
QY 272 CGCTCCACCGCGACAGCT-CTCCCTTCTACCGCTGCGCCGAAACAGGTTGGACGCTTCGTC 330  
DB 1059 CG 1000  
QY 331 GCGCCACCGCTCCAGGAGGAGNACGACGATGACCGCGGCGCATCTCAGCGCGCGCGG 390  
DB 999 GCG 940  
QY 391 CTCACGAGGAGCGCGCGCGCGGACCTCGCCCTGATCTTCGGCGGTTGCTGCGCGCTGCA 450  
DB 939 -NCG 881  
QY 451 GCGCGGCGATCGCGCTCCGGGCTGCTTCGCTGCGGCTTACCATGACGCGCGCGCGCG 510  
DB 880 GCGGNGGCG 821  
QY 511 TGGCTCTGGGACCGCCCAACGCGCGGCTCACCACCGCGCGCTACCGTGTACTACCTG 570

Db 820 GNGNGNGNGCG 761  
QY 571 CACTCGGACGACAAACAA---CGGCGCGCGCGCGCTGGGACCGGACGACGACGACG 626  
Db 760 CGGNGNGCG 701  
QY 627 CGTCGCTTCACGACGACGACGACGACGCTGATGCGCTGCGGCGCGGACTTCCCGCTGTGTC 686  
Db 700 NNGCG 641  
QY 687 CGGTCGCGCGCTGCTGCGCGCACCGCGCAACAGCGGAGTTTCGCGCGCGCGGCTGCTGCG 746  
Db 640 GCG 581  
QY 747 GC-TCGCGACCGACG 805  
Db 580 GCG 521  
QY 806 CGACGACGCGCGGTTTCAGCTTTCACCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 865  
Db 520 GCG 461  
QY 866 GTGCG 891  
Db 460 GCG 435

RESULT 13  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

BM911414 1242 bp mRNA linear EST 12-MAR-2002  
AGENCOURT\_6615160 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5466186  
5', mRNA sequence.  
BM911414  
BM911414.1 GI:19361793  
EST.  
Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 1242)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1966 row: c column: 19  
High quality sequence stop: 350.  
Location/Qualifiers  
1. 1242  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5466186"  
/clone\_lib="NIH\_MGC\_41"  
/tissue\_type="amelanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site:1: xhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

BASE COUNT 47 a 641 c 440 g 55 t 59 others  
ORIGIN

```
Query Match 4.1%; Score 81.2; DB 14; Length 1242;
Best Local Similarity 43.5%; Pred. No. 0.00025;
Matches 373; Conservative 0; Mismatches 480; Indels 5; Gaps 3;

Qy 43 CCGCGGACCGCCCGCGGCGGCTGCTCTCAGCTAGACGGGCGCCCTCTCTGAGGTCTTC 102
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 CCGAGGAGGTTGCGAGTGTGCTGACCTGACGATGATGGGGTCCCGCCGAGGCGTGC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 103 GTCGGGAGCGGTGAGCGACTGCGTCAACCTGTGTCCTGCGGGCGCGGCGTGTGACC 162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 AGGAGGAGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 163 G-CGAGCTCAGACGGCAGCGCAGGAAACCGTGACGTGACCGGATGACGTGAGGC 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 GCGGCGCGCGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 222 GCCAGCGATGCTGACGCGCCCTGACCTGACCTGCGCGCGTCTGCGGCTGACGAGCGTCCAC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 282 CCGACAGCTCTCTTCTTACCGCTGCGGAAACAGGCTGCTGCGGCGCCACCG 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 342 TCCAGGAGGAGAACAGCAATGACGCGCGGCAATCTACGCGCGCGCGCGTGTCTCAGGAGC 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 402 CGGCGCGGAGCACTCGCGCTGTCTGCGGCGTGTGTCGCGCTGACGCGCGGCGATC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 462 CGCTCCGGGCTCGTCTCCGCTGTACACATGACGCGCGCGCGCGCGCGCGCTGCGA 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 694 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 522 CCGCGAGCGCGGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 754 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 582 GAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 814 CNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 640 CACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 700 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 758 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 818 GTTACGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 878 CCAGCGCGCGCGAGATCG 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1114 CNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
```

RESULT 14  
BG852363  
LOCUS 1024034A03.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II  
DEFINITION 1281 bp mRNA linear EST 29-MAY-2001  
Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BG852363  
VERSION BG852363.1 GI:14233547  
KEYWORDS EST.

Chlamydomonas reinhardtii.  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.  
1 (bases 1 to 1281)  
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,  
McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants: project phase 2  
Unpublished (2000)  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1. 1281  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap  
II"  
/note="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP (acetate-containing) medium in the  
light, TAP medium in the dark, HS (minimal) medium in  
ambient levels of CO2 and HS medium bubbled with 5% CO2.  
PolyA mRNA was purified from each sample, pooled and CDNA  
synthesized. The cDNA was directionally cloned into lambda  
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with Exassist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."  
BASE COUNT 30 a 697 c 456 g 15 t 83 others  
ORIGIN

Query Match 4.1%; Score 81.2; DB 12; Length 1281;  
Best Local Similarity 43.1%; Pred. No. 0.00026;  
Matches 418; Conservative 0; Mismatches 547; Indels 5; Gaps 3;  
Qy 9 CCGCGACTTCCTCCGACGACACCGTCCCTACCGCGCGCGCGCGCGCGCGCGCGCG 68  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 314 CCG 373  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 69 CCGCGACTGACG 128  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 374 CCG 433  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 129 GAACCTGTCTCTCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 434 GCG 493  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 189 AACCGTGCACGTGACG 248  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 494 GCG 553  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 249 TCGCGCGCGTCTGGGCTGACGAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 308  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 554 CCG 613  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 309 GAACGAGGTGAGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 614 GCGCG--GGCGGCG 671  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 369 GCGCATCTACGCGCGCGCGTGTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 428  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 672 GCG 730  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
QY 429 CGCGGCTGTGCTCCGCTGCAGCCCGGCGATCCGCTCGGGCTCGCTCCGTCGCCCTCTA 488
Db 731 GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790
QY 489 CCACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548
Db 791 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 850
QY 549 CGCCTACACGCTGTACTACCTCGACCTCGACGAGAACACCGCGCGCGCGCGCGCGCG 608
Db 851 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
QY 609 CCGGAGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668
Db 911 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
QY 669 GCACCTCCCGCGTGTGCTCGCGGTGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCG 728
Db 971 GNCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
QY 729 CGCGCGCGCGCGTGTGCTCGCGTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
Db 1031 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1088
QY 789 GCAGTACCTCTACTGTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848
Db 1089 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
QY 849 CATCTGTCAACACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 908
Db 1149 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1208
QY 909 GTTCCGCGACCCAGATCCACTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968
Db 1209 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1268
QY 969 ACGACTCGCG 978
Db 1269 CCCGAGCGG 1278
```

```
RESULT 15
BO689718
LOCUS BO689718 1051 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8344784 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6247838
5', mRNA sequence.
ACCESSION BO689718
VERSION BO689718.1 GI:21815034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1051)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2386 row: 1 column: 15
High quality sequence start: 2
High quality sequence stop: 283.
Location/Qualifiers
1..1051
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

FEATURES  
source

```
/clone="IMAGE:6247838"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
```

BASE COUNT 63 a 557 c 291 g 69 t 71 others  
ORIGIN

```
Query Match 4.0%; Score 80.8; DB 14; Length 1051;
Best Local Similarity 42.8%; Pred No. 0.00029;
Matches 316; Conservative 0; Mismatches 415; Indels 8; Gaps 1;

...
QY 7 CACCCCGACTTCCCTTCGACGACGCGTCCCTACCGCGCGCGCGCGCGCGCGCGCGCGCG 66
Db 310 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
QY 67 CTCCTCAGCCTAGACGGGCGCGCGTCTCTCGAGGCTTTCGTGCGGGACGGTGAAGGACTGCG 126
Db 370 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429
QY 127 TCGAACCCTGCTCTCTGTTGGGCGCGCGGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 430 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 489
QY 187 GGAACCGGTGACGTGACCGCGGTGACGTGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
Db 490 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 549
QY 247 CCGCGCGCGCGTGTGCGGTGACGAGCGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 550 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
QY 307 CCGAACCAGGTGAGCGTTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 610 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
QY 367 CCGCGCGCGCGTACGCGCGCGCGCGTGTCTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 670 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
QY 427 TTCGCGCGGTGTGTGCGCGCTGCGAGCGCG- - - - -GGCATCCGCTCGCGCGCTCGCTCC 478
Db 730 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
QY 479 GTGCGCGCTTACCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
Db 790 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 849
QY 539 CCACCCAGCGCGCGCGTACGAGTGTACTGCTGACTCGACGACGACGACGCGCGCGCGCGCG 598
Db 850 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 909
QY 599 GCTGGGACCCAGCGAGCAGCGCGCGTGTGCGTTCACGCGCGCGCGCGCGCGCGCGCGCGCG 658
Db 910 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
QY 659 CGCTCGCGCGCGCGTTCGCCGTGTGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 718
Db 970 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029
QY 719 CAGGCTTCGCGCGCGCGCGCG 737
Db 1030 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
```

Search completed: May 10, 2003, 09:52:15

Job time : 2912 secs

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 05:00:26 ; Search time 73 Seconds  
(without alignments)  
8406.304 Million cell updates/sec

Title: US-09-868-328B-2

Perfect score: 2001

Sequence: 1 ggggtgaccccgacttccc.....atcgacacgtctctgctgtt 2001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length  | DB ID | Description       |
|------------|-------|-------------|---------|-------|-------------------|
| C 1        | 101   | 5.0         | 4403765 | 4     | US-09-103-840A-2  |
| C 2        | 97.6  | 4.9         | 2712    | 3     | US-09-025-691-4   |
| C 3        | 94.8  | 4.7         | 4411529 | 4     | US-09-103-840A-1  |
| C 4        | 91.4  | 4.6         | 5392    | 2     | US-08-403-852D-1  |
| C 5        | 91.4  | 4.6         | 5392    | 3     | US-08-510-646B-1  |
| C 6        | 91.4  | 4.6         | 5392    | 4     | US-09-231-818-1   |
| C 7        | 85.6  | 4.3         | 44377   | 2     | US-08-804-227C-7  |
| C 8        | 85.6  | 4.3         | 44377   | 2     | US-08-804-198-1   |
| C 9        | 84    | 4.2         | 8438    | 1     | US-07-945-283-1   |
| C 10       | 83    | 4.1         | 1908    | 3     | US-08-951-742-1   |
| C 11       | 83    | 4.1         | 4403765 | 4     | US-09-103-840A-2  |
| C 12       | 81.4  | 4.1         | 1908    | 1     | US-08-173-508-1   |
| C 13       | 81.4  | 4.1         | 1908    | 2     | US-08-265-310-1   |
| C 14       | 80.4  | 4.0         | 12588   | 2     | US-08-387-942C-1  |
| C 15       | 80.4  | 4.0         | 15872   | 4     | US-09-105-537-1   |
| C 16       | 80    | 4.0         | 1896    | 4     | US-09-311-626B-15 |
| C 17       | 78.6  | 3.9         | 1521    | 1     | US-08-083-948-14  |
| C 18       | 78.6  | 3.9         | 1521    | 1     | US-08-393-785-14  |
| C 19       | 78.6  | 3.9         | 1521    | 1     | US-08-475-694-14  |
| C 20       | 78.6  | 3.9         | 1521    | 1     | US-08-712-057-14  |
| C 21       | 78.6  | 3.9         | 1647    | 1     | US-08-083-948-13  |
| C 22       | 78.6  | 3.9         | 1647    | 1     | US-08-393-785-13  |
| C 23       | 78.6  | 3.9         | 1647    | 1     | US-08-475-694-13  |
| C 24       | 78.6  | 3.9         | 1647    | 1     | US-08-712-057-13  |
| C 25       | 78.6  | 3.9         | 1865    | 1     | US-08-083-948-7   |
| C 26       | 78.6  | 3.9         | 1865    | 1     | US-08-393-785-7   |
| C 27       | 78.6  | 3.9         | 1865    | 1     | US-08-475-694-7   |

|      |      |     |       |   |                   |                    |
|------|------|-----|-------|---|-------------------|--------------------|
| 28   | 78.6 | 3.9 | 1865  | 1 | US-08-712-057-7   | Sequence 7, Appli  |
| 29   | 78.6 | 3.9 | 2182  | 4 | US-09-347-878-35  | Sequence 35, Appli |
| 30   | 78.2 | 3.9 | 11220 | 4 | US-09-105-537-32  | Sequence 32, Appli |
| 31   | 78.2 | 3.9 | 36778 | 4 | US-09-105-537-5   | Sequence 5, Appli  |
| 32   | 78.2 | 3.9 | 38506 | 3 | US-09-320-878-19  | Sequence 19, Appli |
| C 33 | 77.6 | 3.9 | 13613 | 4 | US-09-105-537-3   | Sequence 3, Appli  |
| 34   | 76.4 | 3.8 | 1926  | 4 | US-09-249-585A-4  | Sequence 4, Appli  |
| 35   | 76.4 | 3.8 | 1931  | 2 | US-09-130-114-2   | Sequence 2, Appli  |
| 36   | 75.8 | 3.8 | 1208  | 3 | US-08-403-852D-4  | Sequence 4, Appli  |
| 37   | 75.8 | 3.8 | 1208  | 3 | US-08-510-646B-4  | Sequence 4, Appli  |
| 38   | 75.8 | 3.8 | 1208  | 4 | US-09-231-818-4   | Sequence 4, Appli  |
| 39   | 75.6 | 3.8 | 4689  | 4 | US-09-105-537-34  | Sequence 34, Appli |
| 40   | 75.6 | 3.8 | 13842 | 4 | US-09-105-537-30  | Sequence 30, Appli |
| 41   | 74.4 | 3.7 | 13987 | 2 | US-08-804-227C-13 | Sequence 13, Appli |
| 42   | 73.2 | 3.7 | 1248  | 4 | US-09-105-537-7   | Sequence 7, Appli  |
| C 43 | 72.2 | 3.6 | 5970  | 3 | US-09-320-878-21  | Sequence 21, Appli |
| C 44 | 71.8 | 3.6 | 23673 | 4 | US-09-773-816-1   | Sequence 1, Appli  |
| 45   | 70.8 | 3.5 | 1140  | 3 | US-09-023-173-4   | Sequence 4, Appli  |

ALIGNMENTS

RESULT 1

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 5.0%; Score 101; DB 4; Length 4403765;  
Best Local Similarity 42.3%; Pred. No. 4.1e-10;  
Matches 833; Conservative 0; Mismatches 1115; Indels 23; Gaps 4;

Qy 11 CCGACTTCCCTCGACGACGACGCTCCCTACCGCGGACCGCGCCGCCGGA/TGCTCC 70  
Db 3941672 CGCGCGCGCGCGGACGCGGCTTGCTGTGTCGATATGCGCGGTGCTCCG 3941613

Qy 71 TCAGCTAGAGGGCCCTCTCGAGGTCTTCGTCGGGACGCTGAGCGACTCGCTCGA 130  
Db 3941612 GTGCGCGCGCGCTCCGCGCGCGCGCGGCTCGCGCGGTGCGCGGTACCG 3941553

Qy 131 ACCTGGTCTCTCTGGGGCGCGGCTGTGACCGAGCTCGAGCGGACGCGGCGGAA 190  
Db 3941552 CCGGTCCCGCGATGCGGTGGGGTTATCGGCGCGCGCGCGGCGGCGGCG 3941493

Qy 191 CCGTGCAGTGACGCGGATCGAGTCCGAGGCGGCCAGCGATGCTGAGCCCTGAACCTG 250  
Db 3941492 CCTTGTCCGCGAGTCCGCGCGCGCGGAGCCGTTGTCGCGCGGTACCGCTGTTA 3941433

Qy 251 CCGCGCTTCTGGGTGACGAGCGCTCCACCGCGGAGCTCTCCTTACCGTGCCTCG 310  
Db 3941432 CG 3941373

QY 311 ACCAGGTGGACGTTTGGTTCGGCGCCACCCCTCCACGAGAGGAACACAGCAATAGCGCGG 370

Db 3941372 GTGGCGCGGGCTGGTTCGGCGTTCGGGCGCGCGCGCGCTCCGCGCGTACACGCGCTTGG 3941313

QY 371 CCATCTCACGCCGCGCGTGTCTCCAGGAGACCGCGCGGAGACACTCGCCCTGATCTTCG 430

Db 3941312 CCGCGGCGCGCGACACCGCGGTGGTTCGCGATCATGCGCGCGGTGCGCTTCG 3941253

QY 431 CGGCTGCTGTGCCGCTTCACGCCGGGCATCCGTCGCGGCTCGCTCGCTGCCCTTACC 490

Db 3941252 GTCCGGCGCGCTCCGCGGCGCGCGCGCTCCGCGCGTCCGCGCGTGGCGCGCTCACCG 3941193

QY 491 ACATGACGCCCGCCAGGGCTGGCTTCGCGACCCCAACGCCCGGTTCACCAACCAACGGCG 550

Db 3941192 CCGGTGCCGCGCATCGGTTGGTGGGTTATCGGCGCGCGCGCGCGCGACACCCCGCGCGCG 3941133

QY 551 CTTACCAGCTGTACTACTGCACTCCGACCAAGAACAGGCCCGCGCGGTGGGACCAACG 610

Db 3941132 CCCTGTCCGCGGTGCGCGGCGCGCGCGGAGCGCTTGTGCGCGCGGACACGCTGCTA 3941073

QY 611 CGAGCACGACGACGGGCTCGCCTTCACGACACCAAGCACCGTTCGCTCGCTCGCGCGCG 670

Db 3941072 CCGCGGCGCGCGCGCTTGTCCGCGCGTGCAGCGCGCGCGCGCGCGTGGTGGCGG 3941013

QY 671 ACTTCCCGTGTGTTCGGGTGGCGGTTCGCGACACCGGAGAACAGGCGAGGTTCGCGCG 730

Db 3941012 CCGGCAACCGTGTACCGCGCGCGCGCGCGCGCGCTCCACCGCGGAACCCGGTG 3940953

QY 731 CCGCGCGGCTGTCCGCTCCGACCCAGCCGACCGAGGGGTCCGAAGTACCAAGAGC 790

Db 3940952 CCGCGGCTGGCGCGGCGTGGCGGTGCGCGCGCGCGCGCGCGCTCCGCGCTCACCG 3940893

QY 791 AGTACCTCTACTGTGCACACGCGGGTTTCAGTTTCACCGCGCTCCGCGCGCGCTCA 850

Db 3940892 CCCTTGGCGCGCGCGCGCGACACCGCGCTTGCCTGTGGTG-----CCGATATCGCG 3940839

QY 851 TCGTCAACACGAGGTGGCGCGCGCACCGCGCGCGAGATCGAGAACGCCAGTGGT 910

Db 3940838 CCGGTGCTTCGGTGGCGCGCGTCCGCGCGCGCGCGCTCCGCGCGTCCGCGCGGTG 3940779

QY 911 TCCGGGACCCCAAGATCCACTGGGACACCGCGCGGAGATGGGTCTGCTCATCGGAC 970

Db 3940778 CCGCGGTACACCGCGGTGGCGCGGATGCGCGTGGGTTATCGCGCGCGCGCGCGCAGCA 3940719

QY 971 GACTCGGTATCCGCGCTTCTACACCTCGCGGAACTTCGCGGACTTGGACACTTCGCGCGCA 1030

Db 3940718 CCGCGCGCGCGCTTGTTCGCGCGGTGCGCGCGCGCGCGCGCGGTGGTGGCGCGCG 3940659

QY 1031 ACTTCGACTACCCGAACACCGCCTCGCGCGCATCGAGTGCGCCGACCTGTTCAGATCA 1090

Db 3940658 GCGCCGCGGAGCGTTGGTGGCGCGGACACGCTGCTACCGCGCGCGCGCGCGCTCG 3940599

QY 1091 CCGCAGAGGACGAGACGCGACTGGTGTTCGCGCGCGCGAGTGGAGCGCTACGGCATCG 1150

Db 3940598 CCGGCTCCACCGCGGAACCGCGTTCGCGCGCGGTGGCGCGGAGTGGTTCGCGCTCGCGCGCG 3940539

QY 1151 GCCTCCCATGACTAGCCTACTTGGACAGCACTTGGAGCGGACGACTTCACACGCG 1210

Db 3940538 GCCCGCGCGCTCCGCGCTACCGCGCTTGCGCCCGCGCGCGCGCGCGGTACCGGCGTTACC 3940479

QY 1211 ACGACCTCACCGCGCAATGGCTCGACTGGGCTGGGACTGGTACGCGCGCGCTACCTTGGC 1270

Db 3940478 GCGGCTCCGCTGGCGTGAAGAGCTGCGCGCGCGCACACCGCGCGCGCGCTCACCGCGG 3940419

QY 1271 CATCGATCGAGCGCGCGAGACCAAGCGCCTTCGCCATCGCGGTGGATGAACAACCTGGAAGT 1330

Db 3940418 CTACCGCGCGTTCGCGCTACCGCGCGGTTCGCGCGCGCGCGCGCGGTATCGCGCGCG 3940359

QY 1331 ACGCGCAGCGGCTCCCGACCGACCATCCGAGCGCTACAGGGGCGAGAATCGATCG 1390

Db 3940358 GCGCGCGAGCACCCCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCG 3940299

QY 1391 TCCGCGAGCTCGGCTCGCGCGCAGACGCTGGCGGCTGGTACACCGCTCCTGAGACCGCGCG 1450

Db 3940238 CCAGCGCCCGACAGCAGTGCAGCGCGGCGCCACCGCGCGCGCGGTACCGCCC 3940233  
Qy 1451 TGGCAGCGCTGACCAACTA-----CGTCACCGCGCACACACACTCCCCGACCGGA 1501  
Db 3940238 GTCCGCGCGTACCAACCGCGCGCACCATCGCTGCGGCGACCACTGCGCCCGCCTTGC 3940179  
Qy 1502 CCGTCGACGCGAGCGCGTCTTGCCATGGAACGACGCGCATACGAGATCGAGCTCGACA 1561  
Db 3940178 CCG 3940119  
Qy 1562 TCGCCTGGGACACCGCGACGACGCTCGGCATCTCGGTGGGCGCGCTCCCCGACGGACCC 1621  
Db 3940118 TGTCCGCGCGGTCCCGCGGTAGACACCGGCGCGCGCGCGCGCGCGCGCGCGCG 3940059  
Qy 1622 GGCACAC-----CAACATCGGCAGTACGAGCAGACCTGTACGTGCGACCGAGGACCCCT 1675  
Db 3940058 CCGGCGCGGACAGCAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3939999  
Qy 1676 CCGACCTCGCGGTACTCGCTGCGCCCTTACTCGGAGCGCGCGCGCGCGCGCGCGCG 1735  
Db 3939998 CCGACACCG 3939939  
Qy 1736 CGCGCGCATCGTGCACCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1795  
Db 3939938 GTCCGCGCACCG 3939879  
Qy 1796 ACGCGCGCGCACACGCTGCTCTCCAGCAGGTGCACTTCGCGCGAGGCGCACCGGAATCT 1855  
Db 3939878 GCG 3939819  
Qy 1856 CGCTCTACACG 1915  
Db 3939818 CCCACAATCCGCG--GGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3939761  
Qy 1916 AGCGCATCTAGCGATGACACACACCGCTCACCGAAGCGCGCGCGCGCGCGCG 1966  
Db 3939760 CGGCGCGCGCGGTGCGGAACAGCAGACCGCGCGCGCGCGCGCGCGCGCGCG 3939710

RESULT 2  
US-09-025-691-4  
; Sequence 4, Application US/09025691  
; Patent No. 6069299  
; GENERAL INFORMATION:  
; APPLICANT: Bowman, Gary E.  
; APPLICANT: Bowman, Roxanne M.  
; TITLE OF INVENTION: FUNGUS AND INSECT CONTROL WITH  
; TITLE OF INVENTION: CHITINOLYTIC ENZYMES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/025,691  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/20120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600

## ; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2712 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; US-09-025-691-4

Query Match 4.9%; Score 97.6; DB 3; Length 2712;  
Best Local Similarity 43.9%; Pred. No. 1.4e-09;  
Matches 650; Conservative 0; Mismatches 794; Indels 36; Gaps 4;

QY 466 CCGGGCTCGCTCGCTGCTTACCATGACGCCGCCCGGCGCTGCTTGGACCCC 525  
Db 845 CCGCGCGCTGGCGGCTCGTGGGCTCGGCGTCCCTTCGCGGGTGTGGCGCTC 904  
QY 526 CAACGCCCGGTACACCCAGCGGCGCTACAGTGTACTACCTGCGACTCGGACCAAC 585  
Db 905 GCGGCCCCACCCAGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 964  
QY 586 AACGGCCCCCGGCGCTGGGACACCGCGAGCAGCAGCAGCGCGCTCGGCTTTCACGCACAC 645  
Db 965 CAGGACTGGGCGACGCGCTTCGAGGCGAAGTGACGCTGAAGAACACGGCGCGCGCCC 1024  
QY 646 GGCACCGTGTATCGGCTGGCGCGCGCTTCCCGGTGTGGTCCGGGTGGGCGGTGCTCGGC 705  
Db 1025 CTCAGCGGCTGGACCTTGGAGTGGGACTTCCCGCGCGAACAAGGTGACCTCGGCGCTGG 1084  
QY 706 ACCGCGAACACGGCGGTTTCGGCGCGCGCGCGCGCGCTCGGCTCGGCGACCCAGCCGAC 765  
Db 1085 GACCGCGACGTACCAACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1144  
QY 766 GACGGCTCGCGCAAGTACCGAGGACGACTACTTCTACTGTGCGAGCGCGCGGTTCACG 825  
Db 1145 AGCTCGCGCGCGCGCTCGGTCAGCTTCGGCTTCAACGCGACCGCGCGCGCGCGCG 1204  
QY 826 TTCACCGCTCGCGCGCGCGCTCGTCAACACGAGCGGTGCGCGCGCGCGCGCGCGCG 885  
Db 1205 TCGGGCTGCAAGCTCAACGGCGCGCTTCTGGCGCGCGCGCGCGCGCGCGCGCGCG 1264  
QY 886 GCGGAGATCGAGACGCGAGTGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
Db 1265 CCCACCG 1324  
QY 946 GGAATGCGTCTCGCTACCGACGACTCGGCTACCGCGCGGTTCACACTCGCGCAAC 1005  
Db 1325 TGAAGCG 1384  
QY 1006 CTGGCGGACTGGACACTTCGCGCGCACTTCGACTACCGCAACCGCGCGCGCGCGCG 1062  
Db 1385 AAGGTGCG 1444  
QY 1063 -----ATCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110  
Db 1445 TACTCTACTCGGTTCG 1504  
QY 1111 CACTGGGTGCTCG 1167  
Db 1505 GTCAAGGTACACACCG 1564  
QY 1168 GCCTACTGGACAGCG 1227  
Db 1565 AACCTCGGCTTACTTCACCGAGTGGGCGGTCTACGGCGCGCAACTACCACTCAAAACCTG 1624  
QY 1228 TGGCTCGACTGGGCTGGGAGTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1287  
Db 1625 GTGACCTCCGCGTCCCGCGGAGAGATCACCACATCACTACTCTTCGCGCAAGTCCAG 1684  
QY 1288 GAGACCAAGCGCGCTCGCGCATCGGCTGGATGAACAACCTGGAAGTACGCGCGCGCGCG 1347  
Db 1685 GCGGCAAGTGCACCATCGGTGACAGTTCGCGCGCGCTTACGACAAAGGCGGTACCGCGCG 1744

QY 1348 CCCACGACGATCCGAGCGGCTACACGGGCGAGAACTCGATCGTCCGCGAGCTCGCGCTC 1407  
Db 1745 GAGTCGGTTCGAGCGGCTCGCGGACACCTGGGACCGAGCGCTGCGCGCAACTTCAACAG 1804  
QY 1408 GCCGACAGCTGGCGGCTGGTACACCTCTTGAGCAGCCCCCGTGGCAGCGCTACAGAAC 1467  
Db 1805 CTCGCAAGCTCAAGCGCAAGTACCCGCACATCAAGGTCTCTCTGCTTCTGCGCGGCTGG 1864  
QY 1468 TACGTACCGCCACACACACTCCCGGACCGGCGCTCGACGCGAGCGCC--GTCTCTGC 1525  
Db 1865 ACCTGGTTCGCGGCTTCACCGAGCGCTGAAGAACCGCGCGCTTCGCCAAGTCTCTGC 1924  
QY 1526 CATGGAACGACGCGCATACAGAT-----CGAGCTCGACATCGCCTGG 1569  
Db 1925 CACGACCTGTGCGAGGACCGCGCTGGCGCGAGCGCTCTTCGACGCGATCGACCTCGACTGG 1984  
QY 1570 GACACCGGAGGAGCTGCGCATCTCGTGGCGCGCTCCCGGACGGAACCGCGCACAG 1629  
Db 1985 GAGTACCGGAGCGCTCGGCGCTCAGCTGCGAGCTCGGTCGCGCGCGCTGAAGAAC 2044  
QY 1630 AACATCGGCAAGTACGAGGAGGAGCTGTAGTTCGACGCGAGGACCTCCGACCTCGCGGG 1689  
Db 2045 ATGCTCCAGGAGTGGCGCGCGCTTCGCGACCGACCTGGTTCACCGCGCGCATCACCGCC 2104  
QY 1690 TACTCGCTCGCGCGCTTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1749  
Db 2105 GACGCCAGCTCCGCGGCGAAGCTCGACGCGCGCGCTACGCGCGCGCGCGCGCGCTT 2164  
QY 1750 CACCTGCGCATCTCTGCGACACCGACGAGCTCGAGGCTTCGTCACGCGCGCGCACAC 1809  
Db 2165 GACTGGTACAACTGATGAGTACGACTTCTTCGCGCGCTGGGACAAAGACCGCGCGGAC 2224  
QY 1810 GTGCTCTCCGAGAGTTCACCTTCGCGGAGGCGACGCGGAATCTCGCTCTACACCGAC 1869  
Db 2225 GCGCGCGACCTCGCGCGCTGAACCTCTACAGCGCATCCCAAGGCGGACTTCCACTCGGC 2284  
QY 1870 GCGGCGCGCGCGCATTCACCGGCGATCGTCTCGCGCGAGA 1909  
Db 2285 GCGCGCATCGCAAGCTCAAGGCGGAGGCGTCCCGCGCA 2324

## RESULT 3

US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328

## ; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007 00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 4.7%; Score 94.8; DB 4; Length 4411529;

Best Local Similarity 43.5%; Pred. No. 5.2e-09;

Matches 822; Conservative 0; Mismatches 1027; Indels 42; Gaps 6;

QY 75 CCTAGACGGCGCGCTCTCTCGAGGTCTTCGTGCGGGGACGGTGAGCGGACTCGGTGAACCT 134  
Db 3936368 CCGCGCGAGCGCGCGCTCCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 3936309  
QY 135 GGTCTCTCTGGGCGCGCGGTGTGACCGCGAGCTCTGAGACGCGCGCGCGCGCGCGCT 194

Db 3936308 CGTCACCGCTGGCGCGCTGCTAGCCCGCGTGTGTTAGCGCCGCCCTTTCGCGCTTGGC 3936249  
Qy 195 GCACGTGACCGGATCGAGCTGAGCGGCCAGAGCATGCTACGCGCCCTGAACCTGCGCGC 254  
Db 3936248 CGCGTGGCGCGCGCACCACTTGGCGCGCTCAATGCTGGGTCCCGCTGCGCGCGG 3936189  
Qy 255 CTTTCTGGGTGACAGAGCGCTGCC---ACCGGACAGTCTCTTCTTACCGGTGCCGAA 311  
Db 3936188 CGCGCGCGCGCGCGCGCGCGCGCTGCGCGATCCGCGCTGCGCGGAAACACCG 3936129  
Qy 312 CAAGGTGAGCGCTTCGTGCGGCCACCGCTCCAGGAGAACAGCAATGACGCGCGC 371  
Db 3936128 CGCGAATCCGGCTTTCGCGCGCGCGCGCGATGCGCGCTTGGCGCGCTTACCGCTT 3936069  
Qy 372 CATCTACGCGCGCGCGCTTCAGGAGCGCGCGCGAGCACTCGCGCTTATCTTGGC 431  
Db 3936068 GCGCGCGCGCGCGCGCTGCTGCGCACGCTGCTGCTTCTGCGCGCTACACCGG 3936009  
Qy 432 CGGTGCTGTGCGCGTGCAGCCCGGGCATGCGCTCCGGGCTGCTTCCGTGCGCTTACCA 491  
Db 3936008 CCGCGCGGTGCGCGGTGCGCGCGCGCGCTGATGCGCGCGCGCGCGGTGCGCG 3935949  
Qy 492 CATGACGCGCGCGCGCTGCTGCGACCCCGCGCTCCAGCGCTACACCCAGCGCGC 551  
Db 3935948 CGACCGCGCTTGGCGCGCTTGGCGCGCGCGCGAGCGCTTGGCGTGGGAGAGGCGC 3935889  
Qy 552 CTACAGCTGTACTACTCTCGAGCAGAACACCGCGCGCGCGCTGGGACACCGC 611  
Db 3935888 CGGTGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTGGGACACCGC 3935829  
Qy 612 GAGCAGACCGAGCGCTGCGCTTACGACACCGGACCGCTGATGCGCTGCGCGCGCA 671  
Db 3935828 CCGCGTTCGCGGTGTTGGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935769  
Qy 672 CTTCCCGGTGCTGCGGTGCGGTGCTGCGACCGCGACCGGACGAGGTTCGCGC 731  
Db 3935768 TGGCGCGGACCGCGCATAAAGCGCTGCGCGCTGCGCGCGCTTGGCGCGCGCGCG 3935709  
Qy 732 CGGCGGTGCTGCGCTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791  
Db 3935708 CCGCGCGCTTGGCGCTACCGCTGCGCGCGCTGCTAGCGCGCTGCTTACGCGCG 3935649  
Qy 792 GTACTCTACTGTGACCGAGCGGCTTACGTTTACCGCGCTTGGCGCGCGCGCGCTCAT 851  
Db 3935648 CTTTGGCGCTTGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935592  
Qy 852 CGTCAACACCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 911  
Db 3935591 GGTCCCGCTGCG 3935532  
Qy 912 CCGGACCGCGGATCCACTGGGACACCGCGCGCGGAGAAATGGGTCTGCTCATCGGACG 971  
Db 3935531 CGTGGCGGGAACACCGCGCGCGGAAATCGCGCTTGGCGCGCGCGCGCGCGCGCT 3935472  
Qy 972 ACTGGGTACCGCGGTCTACCTTCCGCAACCTGGCGCGCGCGCGCGCGCGCGCG 1031  
Db 3935471 GCGCGCGCTTACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3935412  
Qy 1032 CTTGACTACCGGAACACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1091  
Db 3935411 TCTGGCGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935352  
Qy 1092 CGGAGACCGGGAACCGCGCTGGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1151  
Db 3935351 CGCGCGCGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935292  
Qy 1152 CTTCCCATGACGTACGCTTACTGACAGGCACTGGGACCGCGCGCGCGCGCGCG 1211  
Db 3935291 CGTCTTGGGAGAGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935232  
Qy 1212 CGACTCACCGCGAATGGCTTCGACTGGGCTGGGACTGGTACGCGCGCGCTACCTGSCC 1271

Db 3935231 CGCGGATACCGCGCTTGGCGCGCGCTTGGCGCGCGCTTGGCGCGCGCGCGCT 3935172  
Qy 1272 ATCGATGACGCGCGCGGAGACCAAGCGCTCGCCATCGCGTGGATGAACAACCTGGAAGTA 1331  
Db 3935171 CGCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3935112  
Qy 1332 CGCGCGCGGAGCTGCCCGACCGAGCGCATCCGACGGCTACACGGGCGAGAACTCATCGT 1391  
Db 3935111 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT-----ACCG 3935071  
Qy 1392 CGCGGAGCTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1451  
Db 3935070 CTGGCGCGCTGGCTAGCGCGCGCGCTGCTTACGGCGCGCTTGGCGCGCTTGG--CGCGCT 3935013  
Qy 1452 GCGAGCGCTGACGAAGTACGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1511  
Db 3935012 GCG 3934953  
Qy 1512 CAGCGCGCTTCTGCGCATGGAACGGAGCGCGCATACGAGATCGAGTTCGATCGCTTGGGA 1571  
Db 3934952 CCG 3934905  
Qy 1572 CACCGCGAGCAAGCTGCGCATCTCGGTGGCGCGCTCCCGCGCGCGCGCGCGCGCG 1631  
Db 3934904 CACCGCGCGCAATCCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934845  
Qy 1632 CATCGCAAGTACGGAGGAGAGCACTGTAGCTGACCGAGGACCGCTTCGCGCGCGCG 1691  
Db 3934844 ATCGCG---GTTGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 3934788  
Qy 1692 CTGCGTCCCGCGCTTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1751  
Db 3934787 GTGCGCGCATGAGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934728  
Qy 1752 CTTGCGCATCTCTGCTGCAACCGAGCGCTGAGGTCTTCTGCAACGCGCGCGCGCG 1811  
Db 3934727 CGCGCGCGCGCTTGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934668  
Qy 1812 GCTCTCCCGAGGTTCCTTTCGCGAGGCGGACACGGGAATCTCGCTCTACACCGCG 1871  
Db 3934667 GCGCGCGCTCAAGCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934608  
Qy 1872 CGCGCGCGCGCATTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1931  
Db 3934607 CGCGCAACCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934548  
Qy 1932 GCACACCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1962  
Db 3934547 CACCG 3934517

## RESULT 4

US-08-403-852D-1

; Sequence 1, Application us/08403852D

; Patent No. 5891695

; GENERAL INFORMATION:

; APPLICANT: Blanc, Veronique

; APPLICANT: Blanche, Francis

; APPLICANT: Crouzet, Joel

; APPLICANT: Jacques, Nathalie

; APPLICANT: Lacroix, Patricia

; APPLICANT: Thibaut, Denis

; APPLICANT: Zagorec, Monique

; APPLICANT: Debussche, Laurent

; APPLICANT: De Crecy-Lagard, Valerie

; TITLE OF INVENTION: Polypeptides Involved In The

; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

; TITLE OF INVENTION: Coding For These Polypeptides And Their Use

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &amp; Dunner

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PC7/FR 93/00923

FILING DATE: 25-SEP-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441

FILING DATE: 25-SEP-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03806, 0054-00000

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5392 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cdna

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: S.pristinae spiralis

US-08-403-852D-1

Query Match 4.6%; Score 91.4; DB 2; Length 5392;  
Best Local Similarity 42.1%; Pred. No. 1.8e-08;  
Matches 652; Conservative 0; Mismatches 891; Indels 6; Gaps 2;

Qy 382 CGCGCGTGTCCAGGAGCGCGCGGAGCACTCGCCCTGATCTTCGGCGGTGCTGTG 441  
Db 2748 CGGACAGATACCCGCGCGGTGCGGCGCGGAGCGCCGGATTCGAGCGCGTCTGTAT 2807  
Qy 442 CGCGCTGACGCGCGGCGCATCGCTCCGGCTCGCTCCGCTGCTACACATGACGCGCC 501  
Db 2808 CGACACCGCGCGCGCGGCTCCAGGCGGTTTCGAGAGGACGAC--GCTGACCGCC 2865  
Qy 502 CCCAGCGCTGCTGTGGACCCCCAACGCGCGGTACACACCCAGCGGCGCTTACAGCTG 561  
Db 2866 GCGCTGGCGCGGTACCGAGCACATCGGCTGTATCACCGCCCGCTCCGCGCGGACCA 2925  
Qy 562 TACTACCTGCACTCGGACGACGAGCGCGCGCGGCTGGGACCGAGGACGACGAC 621  
Db 2926 GCGCGCTACGAGCTGCGGATCACCGCTCGCTCGGACGCTCGCGCGCGGCGCGAC 2985  
Qy 622 GACGCGCTGCGCTTACGACGACGCGACGCTGTGCTGCGCTGCGGCGGCTTCCCGCTG 681  
Db 2986 GCGTGGCTCGGAGCAGGACACACCGCGCGGAGGCGCGGACGCGGCTATCATGAC 3045  
Qy 682 TGGTCCGCGGTGCGGCTGCGGACCGGACGCGGAGGTTTCGCGCGCGGCGCGGCTG 741  
Db 3046 GTCGTCCGCGGCTGTGGACAGCTTCGACGAGGAGGCTTCGTCCAGCGCGCGGAC 3105  
Qy 742 GTCCGCGCTGCGGACCGGACGAGCGGCTGCGGAGTACAGGAGGAGTACCTCTAC 801  
Db 3106 GCGCTGTACTGCGGCTGCGCGCGCTCCACCACTCGACACCGAGGCGGCGGCTTCAC 3165  
Qy 802 TGGTACCGCGGCGGCTTACGTTACGCTTACGCGCGCTGCGCGGCTTACGTTACG 861  
Db 3166 GTGGCGCGCGCGCTTACGTTACGCTTACGCGCGCTGCGCGGCTTACGTTACGTTAC 3225

Qy 862 GACGGTCCGCGCGGACACACCGCGCGGAGATCGAGACGCGGAGTGGTTCGCGGACCCC 921  
Db 3226 GCGCGCGCGCTCGCGCGCGCGCGGACGCTGCTGCTGCTGAGAGCGCGCGCGCGCGCC 3285  
Qy 922 AAGATCCACTGGGACACCGCGCGGAGAAATGGGTCTCGGTCTATCGGAGGACTCGGCTAC 981  
Db 3286 TCGGTGAAGCAGGAGGACCGCGCGGAGGATCTCTGCGCGCTGCGCGCGCGCGCGCGCC 3345  
Qy 982 GCGCGGTTTACACTCTCGCGGAAACCTGCGGAGCTGGAGACTTTCGCGGCAACTTTCGCTAC 1041  
Db 3346 GAACTGCCCGCGGACGCGCGCGGAGCGCTTTCAGGTTGGCGCTCACCGGTTCCGAGCAC 3405  
Qy 1042 CGGAACACGCGCTCGGCGGATCGAGTGGCGGCGGCTGTTTCGAGATCACCGAGACGAC 1101  
Db 3406 CCGTCTCTGGCGCGCTCGCC 3465  
Qy 1102 GGGACACGCACTGGGTGCTCGCGCGGAGATGGAGCGCTACGGAGCGCTCGGCTCGGCTG 1161  
Db 3466 CTGGCGAAGCGCTGGGCTTGGCGCGCGCGCGGAGCGCGCGCTCACCGCGCTGCA 3525  
Qy 1162 ACGTACGCTTACTGGACAGGACCTTGGACGCGCGGAGCACTTCCACGCGGAGCACTAC 1221  
Db 3526 CGACCGCTCGCGCGCTGCTTCTTGGAGA---GTATGTCGCGTGGCTTTCACCTCG 3581  
Qy 1222 CCGCAATGGCTCGACTGGGCTGGAGTGTACGCGGCGGCTGACCTGCGCATTCGATCGAC 1281  
Db 3582 GAGTCCGTGACGAGGCGGCGCGCGGAGCAAGATGCGCGGAGCACTGAGTACGAGCGCTCTC 3641  
Qy 1282 GCGCGCGAGACGCGCTCGGCTGCGGCTGCGGAGTGAACACTGGAAGTACGCGCGGCG 1341  
Db 3642 GACGCGCTGCTGCGGAGGAGCGCGCGCTCACGCGTTCGCGGTCGAGACGCTGATCAC 3701  
Qy 1342 GACGTCCCGCGCGCATCGGAGCGGTACAAACGCGGAGAACTCGATCGTCGCGGAGCTG 1401  
Db 3702 GCGCAGGTCACATCGCGCGGAGGTACACACGAGGCTGAGCGCGCATCGCCCACTG 3761  
Qy 1402 CGGCTCGCGCGGAGCTGGGCTGGTGTACACCTCTCTGAGACCGCGGCTGGAGCGGTG 1461  
Db 3762 GTCCGCGACGAGTCTGCGCTGCGCTGAGTCTGCTGCGGAGGCTGCGGCGCGCGCG 3821  
Qy 1462 ACGACTACGTGTCACCGCGCGGCGGAGTCTCCCGGAGCGGCTGCGGAGGCGCGGCTG 1521  
Db 3822 TCCTGCGGCTGCTCGCTTCTGCTGCGCGGAGTCTCCCGGAGATCGCGCGGCTGCG 3881  
Qy 1522 CTGCGATGGAACGCGCGCATACGAGATCGAGTTCGAGTTCGATCGCTGGAGACCGCG 1581  
Db 3882 AGCGCTAGGAGACCGCGCTGAGGCGGAGGAGGAGCTCGACACGAGCGCGCGCG 3941  
Qy 1582 AACGTGCGCATCTCGGTGGCGGCTCCCGCGGAGAAACCGCGGACACGAGCATCGGCAAG 1641  
Db 3942 GACGAGGCGCTGATGTTGCGGTACGCCACCGAGAGACCGCTGCTGATGCGCGTGGCC 4001  
Qy 1642 TACGAGGAGACCTGTAGTTCGACGCGGAGACCTTCGAGCTCGCGGCTGCTGCTGCG 1701  
Db 4002 ATCGAGCTGCGCGCGCGCTTCTCGCGCGGCTTCACGAGGTCGCGGAGGAGCGGACG 4061  
Qy 1702 CCCTACTCGGAGCGCGCGCGCGCATGACCGCGCGCGGCTGCTGCTGCTGCTGCTGCTG 1761  
Db 4062 CCCTACTCGGCGCGCGCGCGGAGGAGGAGGAGGAGTTCAGTTCAGTTCAGGAGCGG 4121  
Qy 1762 CTCGTCGACACCGCGGCTGAGGTCTTCGTAACGCGCGCGGAGCGGCTGCTTCCCGAG 1821  
Db 4122 GTGGCGCTGGACGCTGCTCTCTCTCCCGAGCGCGCGGAGATGAGCTCGGCTGCG 4181  
Qy 1822 CAGGTTCACCTTCGCGGAGGCGGAGCGGAGATCTCGCTCTACACGAGGCGGCGCGCG 1881  
Db 4182 CTGCTTACCGCGGAGATCCGCGGAGCAGCTGCTGAGGAGCGCTCTGCGCGGAGTTCG 4241  
Qy 1882 CACTTACCGCGGAGTCTCGCGGAGATTCGCGGAGGAGTCTAGGCGCA 1930  
Db 4242 GACGCGATCAAGCTCGAGAGCGGAGCAACTACCGCGCTGCTGGTCAACCGCGA 4290







QY 1222 CCGCAATGCTCGACTGGGGCTGGGACTGGTACGGGGCGGTCACTGGCCATCGATCGAC 1281  
DB 3582 GAGTCGCTGACCGAGGGCCACCCGCAAGATCGCGCCAGATCAGTGACACCGTCTCTC 3641  
QY 1282 GCGCCGGAGACCAAGCGCTCGCCATCGGCTGGATGACAACTGGAAGTACGCCGACGC 1341  
DB 3642 GAGCGCTGCTGCGGAGAGACCCCGCTACCGCTGCGGGTGGAGACCTGTATACCAACC 3701  
QY 1342 GAGTCCCCACCGAGCGATCCGACGGCTACAAAGGCGCAACTCGATCGTCCGGCAGGTG 1401  
DB 3702 GCGCAGGTCCACATCGCGCGGAGGTACCCACCAAGGCTAGCGCCCATCGCCCACTG 3761  
QY 1402 CGGCTGGCCGACACCTGGCGGCTGGTACACCTCTTACGACACCCCGTGGCAGCGTG 1461  
DB 3762 GTCCGCGACACGATCTGCGCATCGCTAGGACTCGTCCGCAAGGCTTTCGACGGCGCC 3821  
QY 1462 AGCAACTAGTACCGCCACACGACACTCCCGGACCGGAGCGTGGAGCGGCGCGCTC 1521  
DB 3822 TCCTGGGGCTCTCGCTTCCATCGCGCGGCTAGTCCCGGACATCGCCAGGGCGTGCAC 3881  
QY 1522 CTGCCATGAACGAGCGGCGATACGAGATCGAGCTCGACATCGCTGGGACACCGCGAGC 1581  
DB 3882 AGCGCTACGACACCGCGTTCAGGGCGAGGAGGAGCTCGACGAGCGCGCGCGC 3941  
QY 1582 AACGTGGGATCTCGGTGGCGCTCCCGGACGGAACCGGACACCAACATCGGCAAG 1641  
DB 3942 GACCAAGGCGCTGATCTCGGTACGCCACCGACGAGACCCCTCGCTGATGCGCGTGC 4001  
QY 1642 TACGAGCAGACCTGACTGACGACGAGGACCTCCGACCTCGCGGCTACTCGTCTGCC 1701  
DB 4002 ATGAGCTGCCACCGCTCTCGCGCGGCTACCGAGTCCGAGGTCGGAAGGACGCGACCTG 4061  
QY 1702 CCTACTCGGAGCGCGCGCCCGATCGACCCCGCGCGGCGGATCGCTGACCTGGCGATC 1761  
DB 4062 CCTACTCGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4121  
QY 1762 CTCGTGACACCGAGGCTGAGTCTTCTGTCAGCGCGGCGGACACCGTGTCTCCAG 1821  
DB 4122 GTGCGCTTGACACCGTGTCTCTCTCCAGCAGCGCGGAGACGAGTCTGCGGTCC 4181  
QY 1822 CAGGTCCACTTCGCGGAGGCGGACACGGAATCTCGCTTACACGAGCGGCGCGCGCA 1881  
DB 4182 CTGCTACCGCGGATCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4241  
QY 1882 CACTTACCGGATCGTCTCGCGGAGATGGCCAGGCGATAGGCGA 1930  
DB 4242 GACGGCATCAAGCTCGAGAGGAGCAACTAGCGCGTGTGGTCAACCCGA 4290

RESULT 7  
US-08-804-227C-7  
; Sequence 7, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kustoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 31329..36071  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 36155..41830  
; US-08-804-227C-7  
  
Query Match 4.3%; Score 85.6; DB 2; Length 44377;  
Best Local Similarity 42.1%; Pred. No. 2e-07;  
Matches 727; Conservative 0; Mismatches 969; Indels 30; Gaps 3;  
  
QY 69 CTTACGCTAGACGGGCCCCCTCTCGAGGTCTTCTCGGGGACGGTGGAGCGACTGCGCTC 128  
DB 26238 CCGCACCGGTTTCGAGCAGCGGGCGCGCTCTCGCGGAGAACTTCACGGAGCTGATCGC 26297  
QY 129 GAACCTGTCTCTCGGGGCGGGGTGTGACCGGAGGCTCGAGAGCGGCGGCGGCGGCGG 188  
DB 26298 GCGCTCTGACGACCTCTGTCGAAGCGCGGCGGCGGCTCTCTCTGCGGGGCGGCGGCG 26357  
QY 189 AACCTGTGACGCTGACCGGATCGAGCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 248  
DB 26358 CACCTCCGACGAGTCTGTTCTCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 26417  
QY 249 TGCCCGCGTCTGAGGCTGACGAGCGCTCCACCGCGGCGGCGGCGGCGGCGGCGGCGG 308  
DB 26418 CGACGGGCTGCTGGCGCGCTCCAGCGGCTCTCTCTCTGAGAGCGGCGGCGGCGGCTG 26477  
QY 309 GAACGAGGTGGACGCTTCTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 368  
DB 26478 CGACCTCGGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 26537  
QY 369 GCGCATCTCAGCGCGCGGCGGCTCCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 428  
DB 26538 CCGCGCGGCTCGCTCGACCGGCTCGAGCTGTCGAGCGGCGGCGGCGGCGGCGGCGG 26597  
QY 429 CGGCGGCTGTCGCGGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 488  
DB 26598 CTGCGTCCGCGAGAGCTGGCGCTTCTGCGGCTTCGAACCGGCGGCGGCGGCGGCGG 26657  
QY 489 CCACATGACCGCGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 548  
DB 26658 CCAGGCGAGATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 26717  
QY 549 GCGCTTACGAGTGTACTACTCTGACCTCGGAGCAGAACGCGGCGGCGGCGGCGGCGGCG 608  
DB 26718 CATGTCGCCCTCGCGCAGCGGCGGCTGCTGCGGCTGGCGCGGCGGCGGCGGCGGCGG 26777





Db 26238 CCGACACCGGTTGAGACAGCGGGCCCGCTGCTGCGGAGAACTTCCACGGAGTGATCCG 26297  
QY 129 GAACCTGGTCTCTCTGGGGCGGGCTGTGACCGGAGACCTCGAGACGCGACGGCCAGG 188  
Db 26298 GGGCTCGAGGACCTGCTCGAAGGCGGACCGACCGCTGCTGCGGGGACCGCGCG 26357  
QY 189 AACCTGACAGTACCGCGATGACGTGAGAGGCGCCACCGGATGTCAGCGCCCTGAACC 248  
Db 26358 CACCTCGGACAGTGTGCTGCTTCCCGGCGCCAGGCTCGCAGTGGCGCGAGATGC 26417  
QY 249 TGGCGCGCTTCTGGGCTGACGAGCGCTCCACCGGACAGCTCTCTCTTACGGTGCC 308  
Db 26418 CGACGGGCTGCTGGCCCGCTCCAGGGCTCCGGCTCTCTTCCGAGACCGCCCGGCTG 26477  
QY 309 GAACAGGTTGACGCTTCTGTCGCGCCACCGCTCCACGAGAGAACACGATGACGCC 368  
Db 26478 CGACCTCGGCTCCGGCCCGACCTCGGCTGGTGGCTGACGATCTCGCGCGGAACC 26537  
QY 369 GGGCATCTACCGCGGCGGCTGCTCAGGAGCGCGCGGAGACACTCGCCCTGATCTT 428  
Db 26538 CGGCGCGGCTCGCTCGACCGGCTGACGTGTGACGCGGCTGCTGTTCCACCATGATGT 26597  
QY 429 CGGCGGCTGCTGCGGCTGACGCGCGGACATCCGCTCGGGCTGCTCTCTGCGGCTTA 488  
Db 26598 CTCGCTCGGAGAGCTGGGCTTCTGCTGGGCTGACGCGCGGCTGCTGCTGCTC 26657  
QY 489 CCACATGACCGCGCGGAGCGGCTGCTGCGACCGCCCAACCGCGGTCACACCGACGG 548  
Db 26658 CCAGGCGAGATCGCGCGGCTACGTCTCGCGGCGCTGACGCTGACGACGCGCGG 26717  
QY 549 CGCTACGAGCTGTACTACTGCTCCGACCGACGACGACGCGCGCGGCTGGGACCA 608  
Db 26718 CATCTGCGGCTGCGGACGAGCGGCTGCGGCTGCGGCGGACGAGCGGATGTCG 26777  
QY 609 CGGAGACGACGCGGCTGCGCTTACGACGACGACGACGCTGATCCGCTCGGCG 668  
Db 26778 GGTGACCTGTCGACGAGGAGCTGGGCTCCCGCTGAGGCTGAGGACCGGCTCG 26837  
QY 669 CGACTTCCCGTGTGCTGCGGCTGCGGCTGCGGACCGGAGACGAGGCTT--- 725  
Db 26838 CGTGGCGCGCTCAAGGCGCGGAGACCTGCGCGCTCTCCGGGACCGGACCGCTGC 26897  
QY 726 -----CGGCGCGGCGGCTGCTGCGGCTGCGGCTGCGGACCGGACCGGCTG 773  
Db 26898 GAGCTGTGCGCGGACTGCTGCGGAGGCGGCTGACGCGCGGCTGCGGCGGCTGCA 26957  
QY 774 CGCAAGTACGAGGAGCTACTCTACTGCTGACGCGGCGGCTTCACTTCAACGC 833  
Db 26958 CACCGCGGCGACTCGCGGAGTGCACGCTGGAGGCGGCTGCGGAGGCTGCTGC 27017  
QY 834 CTGCGCGGCGGCTGCTCAACACGAGCGGCTGCGGCGGCGGACCGCGCGGAGAT 893  
Db 27018 GCGGCTGCGGCGGCGGCTGCGGACCTGCTTACTGCGGCTGACGCGGAGGACTGAT 27077  
QY 894 CGAGAACGCGGAGTGTGCGGACCGGCGGCTGCGGACCGGCGGCGGAGATG 953  
Db 27078 CGACACCGCGGAGCTGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27137  
QY 954 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1013  
Db 27138 CGAGCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27197  
QY 1014 CTGGGACTTCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1073  
Db 27198 GCACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27257  
QY 1074 CGACCTGCTGAGATCACCGGAGACGCGGACGCGGCTGCTGCTGCTGCTGCTGCTGCT 1133  
Db 27258 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27317  
QY 1134 GCAGCGCTACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1193  
Db 27318 CCGCGGCTACACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27377

QY 1194 CGAGCAGTTTCCAGCGCGAGACCTCACCCCGCAATGGCTGAGCTGGGGCTGGGACTGGTA 1253  
Db 27378 CCAGGTGGAACCTGCCACGCTACCTCCCTTCAGCGCGAGCGCTACTGTACAGCCCGGCA 27437  
QY 1254 CCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1313  
Db 27438 CCGCGGCTGACGAGCCCGGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27497  
QY 1314 GATGAACAACTGGAAGTACGCGCGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1373  
Db 27498 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27557  
QY 1374 CCGGCGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1433  
Db 27558 CACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27617  
QY 1434 CTTCTGAGACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1493  
Db 27618 CG-----CGACCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27668  
QY 1494 CGACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1553  
Db 27669 GAGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27728  
QY 1554 GCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613  
Db 27729 GAACGTGCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27788  
QY 1614 CGGAACCGGCGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1673  
Db 27789 GGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27842  
QY 1674 CTCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1733  
Db 27843 GCGCGACGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 17902  
QY 1734 CGGCGCGGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1779  
Db 27903 CGGCGACTTCTACGAGCGCTTCTCGGAACTCGGTTACGGCTACGGC 27948

## RESULT 9

US-07-945-283-1/c  
; Sequence 1, Application US/07945283  
; Patent No. 5352596  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Andrew K.  
; APPLICANT: Wesley, Ronald D.  
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
; TITLE OF INVENTION: Involving The EP0 and LLT Genes  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis P. Ribando  
; STREET: 1815 No. 5352596th University Street  
; CITY: Peoria  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 61604  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/945,283  
; FILING DATE: 19920911  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ribando, Curtis P  
; REGISTRATION NUMBER: 27976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 309-685-4011 ext.513

TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8438 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudorabies virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 622..6495  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1099, "g")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1267, "t")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1381, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1566, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(7010, "g")  
US-07-945-283-1

Query Match 4.2%; Score 84; DB 1; Length 8438;  
Best Local Similarity 43.08; Pred. No. 3.7e-07;  
Matches 640; Conservative 0; Mismatches 830; Indels 18; Gaps 4;

Qy 329 TCGCGCCACCCGTCACGAGAGAACACCAATGACGCCGGGCATCTACGCCGCGCGC 388  
Dy 5549 TCG 5490  
Qy 389 TGCTCCAGGAGCG 448  
Dy 5489 CCTCCTCGCGCTCCG 5430  
Qy 449 CAGCG 508  
Dy 5429 CCG 5370  
Qy 509 GCTGGCTCGGACCCCG 568  
Dy 5369 ACGGGCG 5310  
Qy 569 TGCACTCGGACAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628  
Dy 5309 CCGACCTTCG 5250  
Qy 629 TCGCGCTTCACGACCG 688  
Dy 5249 TCTGGAGGAGGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5190  
Qy 689 GGTGGCGGCTCGGCG 748  
Dy 5189 TGCGCGCTTCATCCGCGGAGATGGGGACTCGAGGAGCAGCAGGAGCGCGCGCG 5130  
Qy 749 TCGCGACCCCG 808  
Dy 5129 TCATCTACAGCG 5070  
Qy 809 CCGAGCGCGGTTACGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868  
Dy 5069 CCG 5010  
Qy 869 GCG 928

## RESULT 10

US-08-951-742-1  
; Sequence 1, Application us/08951742  
; Patent No. 6127144  
; GENERAL INFORMATION:  
; APPLICANT: Bartfield, Daniel  
; APPLICANT: Michael J. Butler  
; APPLICANT: Dany Hadary  
; APPLICANT: David Jenish

Db 5009 GCTCCTTCATCACCGGAGCGTGACCCGCGCTCGCGACATCGGGAGCCCATGGCGG 4950  
Qy 929 ACTGGGACACCGCGCGGGAATGGGTCTCGGTCTATCGGACGACTGGGTACCGCGGT 988  
Dy 4949 CCCAGGACCGCGCTCTGGGCCCTCGCGACGCGGTGAGCGGTGGCCATGAGCGCGCG 4890  
Qy 989 TCTA---CACCTCGCGGAACCTGCGGACCTGCGGACACTTCGCGCAACTTCGACTAC 1045  
Dy 4889 ACGATCGCACCCAGAACCTTCATCTCGAGAGCCTTCGCGGCGCTACCGCGACATGG 4830  
Qy 1046 ACCACGCGCTCGCGGCGATCGAGTGCCTCGGACCTTCGAGATCACCGAGACAGCGGA 1105  
Dy 4829 CCTACCGGCG 4770  
Qy 1106 CACGCCACTGGTCTCG 1165  
Dy 4769 CCGCGCTCGCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4710  
Qy 1166 ACGCCT---ACTGGACAGGCGACCTTGGGACGCGGAGCAGTTCCACGCGGACGACCTC 1222  
Dy 4709 ACGCTGGCTCTGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4650  
Qy 1223 CGCAATGGCTCGACTGGGGCTGGGACTGGTACGCGCGCGCTACCTGGCCATTCGATCG 1282  
Dy 4649 TGCG 4590  
Qy 1283 CGCGCGGAGACCAAGCGC-----TCGCCATCGGCTGGATGAACAACCTGGAGTAC 1333  
Dy 4589 CCGCGCTGCTCGCGCTCGCGGACGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 4530  
Qy 1334 CCGCACGCGAGCTCCCGACCGACGATCCGAGCGCTACACGCGGCGCAACTTCGATCG 1393  
Dy 4529 CCGCGGACTCTGCTCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4470  
Qy 1394 GCGAGCTCGGCTCG 1453  
Dy 4469 ACGGGGCGCG---CGCGCGCGGAGCGCGCTGGACCTACGCGCGCGCGCGCGCGCG 4413  
Qy 1454 CAGCGCTGAGAACTAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1513  
Dy 4412 CCAAGTGGCG 4353  
Qy 1514 GCGCGCTCTGCCATGGAACGAGCGCGCATACGAGATCGAGCTCGACATCGCTTGGGACA 1573  
Dy 4352 GCGCGCGCTGCG 4293  
Qy 1574 CCGCGAGACGCTCGGATCTCGGTGGCGCGCTCCCGCGGAGGAAACCGCGGACAGCA 1633  
Dy 4292 CCGCGGCG 4233  
Qy 1634 TCGCAAGTACGAGGAGACCTGTACGTGACGCGGAGGACCTCCGACCTCGCGGGTACT 1693  
Dy 4232 CCAAGTCCGGTCCAGACCAAGTCCAGCTTCGCGGACCAAGTTCGCGGCTCTCTCG 4173  
Qy 1694 CGCTCGCGCGCGCTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1753  
Dy 4172 CCGGCTACGCCAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4113  
Qy 1754 TCGGCATCTCTGTCGACCGCGAGCGCTGAGGCTTCCTGTCACGCGCG 1801  
Dy 4112 AGCG 4065

```

1  APPLICANT: Tim Krieger
2  APPLICANT: Lawrence T. Malek
3  APPLICANT: Gisela Soostmeyer
4  APPLICANT: Eva Walczyk
5  APPLICANT: Phyllis Knygman
6  APPLICANT: Sheila Garven
7  TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN
8  TITLE OF INVENTION: BACTERIAL HOST CELLS
9  NUMBER OF SEQUENCES: 25
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Foley & Lardner
12 STREET: 3000 K Street, N.W., Suite 500
13 CITY: Washington
14 STATE: D.C.
15 COUNTRY: USA
16 ZIP: 20007-5109
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: PatentIn Release #1.0, Version #1.30
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/951,742
24 FILING DATE: 16-OCT-1997
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Bent, Stephen A.
27 REGISTRATION NUMBER: 29,768
28 REFERENCE/DOCKET NUMBER: 0189740/0140
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (202)672-5300
31 TELEFAX: (202)672-5399
32 TELEX: 904136
33 INFORMATION FOR SEQ ID NO: 1:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 1908 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39 MOLECULE TYPE: cDNA
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 146..1756
43 FEATURE:
44 NAME/KEY: misc.feature
45 LOCATION: 146..148
46 OTHER INFORMATION: /product= "Met at position -39
47 OTHER INFORMATION: represents fMet"
48 FEATURE:
49 NAME/KEY: sig_peptide
50 LOCATION: 146..262
51 FEATURE:
52 NAME/KEY: mat_peptide
53 LOCATION: 263..1756
54 US-08-951-742-1

```

|    | Query Match           | 4.1%;  | Score 83;          | DB 3;           | Length 1908;           |
|----|-----------------------|--|--------------------|-----------------|------------------------|
|    | Best Local Similarity | 43.2%;   | Pred. No. 5.4e-07; |                 |                        |
|    | Matches 560;          | Conservative   | 0;                 | Mismatches 720; | Indels . 15; Gaps . 3; |
| QY | 610                   | GCAGACGACGACGCGCTCCGCTTACGACACACGACCGCTGATCCGCTGCGGCC      | 669                |                 |                        |
| Db | 101                   | CGAGGATCCCGTACTTGTCCGAACACGTACGGGAGGGCCACATGAGGAAGACAGC    | 160                |                 |                        |
| QY | 670                   | GACTTCCCGGTGTCGGGTCCGGCTCGTCGGCACCGCAACACGGCAGGGTTCGGC     | 729                |                 |                        |
| Db | 161                   | ATACGGCGGAGGGACCGGCTTCGGACACGCCGGAGCACTGGTCACGCCACGCTGATC  | 220                |                 |                        |
| QY | 730                   | GC CGCGCGCGGTCTCGCGTCCGACCCAGCCGACGCGCGTCCGCAAGTACCGAGG    | 789                |                 |                        |
| Db | 221                   | GC CGGCGCGCGTCTCGGACACCGCGCAGCGCCGCCCGCGGCGAGCGGCACGGCACGG | 280                |                 |                        |
| QY | 790                   | CAGTACCTCTACTGGTCGACGCGGGGGTTCAGTTACCGGCCCTTCCCGACCCCGTC   | 849                |                 |                        |

|    |      |                               |                                    |      |
|----|------|-------------------------------|------------------------------------|------|
| Db | 281  | CGAGCTGGGACCGGAGGCGCGGGTGC    | CGCATCGCGCGCGCGCGCGCGCGCGGCG       | 340  |
| QY | 850  | ATTCGTCAACACGACGGTTCGCGCGCCAC | CACACGCGCGCGGAGATCGAGAACCGCGAGTGG  | 909  |
| Db | 341  | GGCATCGACTGGGAGACTGCGCAGCGACT | TGGAACC-----TGCCCAA:CCCATCCAG      | 394  |
| QY | 910  | TTCCGGGACCCCAAGATCACATGGGACAC | CGCGCGGGAGAAATGGTCTGCGTCATCGGA     | 969  |
| Db | 395  | TGCGGCTACGTACGTCGCGATGGACTAC  | GGCAAGCCGTACGGCAAGCAGATACAGGCTC    | 454  |
| QY | 970  | CGACTGCGGTACGCGCGCTTACACCTC   | GC                                 | 1029 |
| Db | 455  | GCCGTGCGACCGCATCGGCAACACGGAA  | CACAGAGGAGCGCAGCGCGCGCCCTCATCTAC   | 514  |
| QY | 1030 | AACTTCGACTACCCGAACACACGCCCTC  | GC                                 | 1089 |
| Db | 515  | AACCCGGGGTTCGCGCGGCTTCGGCC    | TTCGGCTTCGCGCGCGCTCAGCAACAGAGC     | 574  |
| QY | 1090 | ACCGCAGACGCGGACACGCGCACTGGT   | GCTCGCGCGCAGCATGGACGCCCTACGGCATC   | 1149 |
| Db | 575  | CGGGTCTGSGCCAAACAGGCCAAGGCT   | TACGACTTCGTCGGCTTCGACCCGCGCGCTC    | 634  |
| QY | 1150 | GGCCTCCCATCAGTACGCCCTACTTGG   | ACAGGCACTCTGGAGCGCGCAGCATTTCCAGGCC | 1209 |
| Db | 635  | GGCCACTCC--GCGCCCATCTCTCTG    | CTGCTGACCCGCGAGGAGTTCGTAAGGCACCAAG | 691  |
| QY | 1210 | GACGACCTCACCCCAATGGCTCGACT    | GGGGCTGGGACTGTACGCGCGCTCACCCTGG    | 1269 |
| Db | 692  | GCCGACCGCTCCCGGCTCCGAGCGCG    | ACAGCGCGCCAGGCAAGCTCGCCCGCGAG      | 751  |
| QY | 1270 | CCATCGATCGAGCGCGCCGACCAAGC    | CGCTTCGCATCGCTGGATGAACAACTTGGAA    | 1329 |
| Db | 752  | TACGCGGAGGGTGTTCGAGCGAGCG     | CGAGATGCTCCCGCACATGACCCGCGGAAC     | 811  |
| QY | 1330 | TACGCGCAGCGGAGTCCCGACGACG     | ATCGGAGGCTACAGCGGCGAGAACTCGATC     | 1389 |
| Db | 812  | ACCGCGCGGACTCGAGGCTCATCGC     | CGCGCGCTTCGGCGAAGAAGTCAACT-----    | 866  |
| QY | 1390 | GTCCCGGAGCTCGGCTCGCCGACAG     | CGCTCGCGGTGTACACCTCTCTAGCACCCCC    | 1449 |
| Db | 867  | -ACCTCGGGTCTCTACGGCACTTAC     | TCTGCGCGGCTCTACGGCACTCTTCCCGGAC    | 925  |
| QY | 1450 | GTGCGAGGCTGACGAATCTACGTCA     | CGCGCACCACTCCCGACCGGACCGCTCGAC     | 1509 |
| Db | 926  | CAGCTCCCGCATGGTGTGCACAGC      | GTCTCAACCCGTCCGCGACAAGATCTGGTAC    | 985  |
| QY | 1510 | GGCAGCGGCTCTGCATGGAAACG       | CGCGCATACGAGATCGAGCTCGACATCGCCTG   | 1569 |
| Db | 986  | CAGGCAACTGGACCGAGACCTCGC      | CTTCGAGGGCGCTGGAAGGACTTGGCAGGACT   | 1045 |
| QY | 1570 | GACACCGGACAACTCGGCATCTCG      | TGGCGGCTCCCGCGACGGAACCGCGCACAG     | 1629 |
| Db | 1046 | GTGCGCGGAGACCGCGCTTACC        | ACTTCGGGACACCCGCGCGAGTCCAGGACCAG   | 1105 |
| QY | 1630 | AACATCGGCAAGTACGAGCAGAC       | CTGTACTCGACCGGACCTTCGACCTCGCGGG    | 1689 |
| Db | 1106 | TGGCTGAAGCTCGCGCGCGCCCG       | CGAGAACGCTGGCGGCTCTCGTGACCGGCG     | 1165 |
| QY | 1690 | TACTCGTGGCCCTACTTCGGAGCG      | CGCGCGCCATCGACCCCGCGCGCGATCCGTTG   | 1749 |
| Db | 1166 | GAGCTGATCTCTTCTTCAGAGC        | CGCGCTACTACGACTTCGCGCTGGCGCGCAG    | 1225 |
| QY | 1750 | CACCTGGCATCTCTGTGCACAC        | CAGAGCTTCGTTCTTCAACGCGCGGCACAC     | 1809 |
| Db | 1226 | GAGATCTTACGAAAGTACGTTC        | CGCGCGGACACCCAGCGCTCTGTGACCGCGC    | 1285 |
| QY | 1810 | GTGCTCTCCACAGGTCCACTTTC       | CGCGAGGCGACACGGGAATCTCGCTTACACCG   | 1869 |
| Db | 1286 | GACCTGTCCGACACCGCGGCA         | ACGCCCTCCCGCGAGACGGCAACGCGCTCTAC   | 1345 |
| QY | 1870 | GGCGGCCCGCACACTTCA            | CGGCAATCTGCTGCTTCG                 | 1904 |
| Db | 1346 | GTGAGTGCACCGCAAGTGG           | CGCGCGCAACTG                       | 1380 |

## RESULT 11

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 4.1%; Score 83; DB 4; Length 4403765;  
Best Local Similarity 42.6%; Pred. No. 6.6e-07;  
Matches 774; Conservative 1; Mismatches 1029; Indels 12; Gaps 4;

QY 86 CCTCTCGAGGTCTTCGTCGGGACGCTGAGGCGACTGGGTGCGAAGCTGGTCTCTCTGG 145  
Db 1631220 CCTCTGTTGGTCGAGGAGCGCCCTTGGCGCGGTCTCGCGCGCGCCAGTCCCGCG 1631279

QY 146 GGGCGCGGTGTACCGCGAGCTCTGAGAGCGGACGCGGCGAGGACGTCAGTACCG 205  
Db 1631280 ATGCCACCGCTTGCCTGGTGGCGCGGTGGCGCGGTACCGCGCTTACCGCGGTGCGCG 1631339

QY 206 CGATCGACGTCGAGGCGCGCCGAGCGATGCTGACGCGCCCTGAACCTGCGCGCGTCTCGGCT 265  
Db 1631340 CCGGGCGCGCTTTCGCGCGTGGCAGCGCGTGGCGCGGTGCGCGCGGTGCGCGCG 1631399

QY 266 GAGCAGCGCTCCACCCGCGAGCTCTCTTACCGCTGCGCGAAGCAGCGGTGAGCGCT 325  
Db 1631400 TCGTCMGATGCTGCGCGCGCGCCCGCTTCCCTGCGGTGAGCGCGGTGCGCGCGG 1631459

QY 326 TCGTCGGCGCCACCGCTCCAGAGAGAACAGCAATGACGCGCGCGCATCTCAGCGCGG 385  
Db 1631460 CCGCGCGCACCGCGGTGCGCGCGTGGCGCGGTGCGCGCGTGGCGCGCGCGCGCG 1631519

QY 386 CCGTGTCTCAGGAGCGCGCGCGGAGCACTCGCCCTGATCTTCGGCGGTGCTGTCGCG 445  
Db 1631520 GCGAGGCGCCCTGCTCCCTGCGCGGTGCGCGCGTGGCGCGGTGCGCGGAGCAGCGCG 1631579

QY 446 CTGACGCGCGGATCGCTCCGCGGTGCTGCGTCCGCTGCTACCACTACGCGCGCCCA 505  
Db 1631580 TCGCGCGCGCGCCACCGGACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1631639

QY 506 GCGGCTGGCTCTGGAGCCCGCCAGCGCGGTGCTACACCGCGCGCGCTTACAGCTGTACT 565  
Db 1631640 CCGCGCGCGCGCGGTGCGCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGGTGCG 1631699

QY 566 ACCTGCACTCCGACCAAGAACGCGCGCGCGGTGGGACCGCGGAGCAGCAGCCAGC 625  
Db 1631700 GCGGACCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631756

QY 626 GCGTCTGCTTACGACCGCGCGCGGTGATGCGGTGCGCGCGCGCGCGCGCGCGCGCG 685  
Db 1631757 AGCGCGCGGTGCGCGCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631816

QY 686 CCGGGTGGCGGTCTGTCGCGACCGCGAAGCAGGAGGTTTCGGCGCGCGCGGTCTGTCG 745

Db 1631817 CCGGCGCCAC---CGTTGCGCGCGGGTTACCGTTTGGCCCTTTTCACGAGGTGGTG 1631873

QY 746 CGCTCGCGACCGCGAGCGCGCTCGCAAGTACAGGAGCAGTACTCTACTGGT 805  
Db 1631874 GCGTTGGCACTATGCCACCAAGCGCGGTGCGCGCGCGCGGTGCGCGCGCGTG 1631933

QY 806 CGACCGAGCGCGGTTCACGTTTACCGCCCTGCGCGCGCGCTCATCGTCAACACCGAGC 865  
Db 1631934 CCGGCGCTGCGCGGTTCGCGCACTTGCCTGCGCGCGGTGCGCGCGCGCGCTGG 1631993

QY 866 GTGCGCGCGCGACACCGCGCGCGAGATCGAGAACCGCGAGTGGTTCGCGACCGCCAGA 925  
Db 1631994 GAGTTGCGCGCGGTTCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTG 1632053

QY 926 TCCACTGGGACACCGCGCGCGGAGATGGTCTCGCTCATCGGACGACTGCGGTACGCGC 985  
Db 1632054 GCGCCCTTGGCGCTGGGTAGAGCGCGCGCGCGCTACCGCCCTGCGCGCGCGCG 1632113

QY 986 CGTTCTACACTCGCGAAGCTGCGGACTGGACACTTCGC---CGCAACTTCGACTACC 1042  
Db 1632114 GGGTTACCGCGCTTACCGCGGTACCGCGCTTACCATCGCGAAGCGAAGTTGCGG 1632173

QY 1043 CGAACACCGCGCTGCGCGCGCATGAGTGGCGCGCGCGCTTTCGAGATCACCGAGACG 1102  
Db 1632174 GCGCGCTTTCGCGCGGTACCGCGCGCGCGCGCGCGCGCTTTCGCGCGCGCGCG 1632233

QY 1103 GGACACGCACTGGGTCTGCGCGCGAGCATGGAGCGCTACGGCGCTGCGCTCCCATGA 1162  
Db 1632234 ACACCTGATTCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTG 1632293

QY 1163 CGTACGCTTACTGGACAGGACCTGGGACGCGCGAGAGTTCACGCGCGAGCACTCAC 1222  
Db 1632294 CCGCGCTTACCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTG 1632353

QY 1223 CGCAATGCTCGACTGGGCTGGGACTGTACGGGCGCTGCTGCGCATGCTGCGCATG 1282  
Db 1632354 GCGGTGGTGGCGTGGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTG 1632413

QY 1283 CGCGCGAGACGAGCGCTCGCATCGCTGGATGAACAACTGGAAGTACGCGCGAGCG 1342  
Db 1632414 CCGCGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCG 1632470

QY 1343 ACGTCCCGACGAGCATTCGCGCGGTACAAAGGCGAGAACTCGATCGTCCGCGAGTGC 1402  
Db 1632471 CCGCGCGCGCGCGCGCGCGGTTCGCGCGCTGCGCGCGCTGCGCGCGCGCGGTG 1632530

QY 1403 GGTGCGCGCGCGCGCGCGGTGCGCGCGGTGCGCGCGCTGCGCGCGCGCGCGCGCG 1462  
Db 1632531 GCGGAGGAAACCGCGCGACCGCGCATGCGCTCCCGCGCTTTCGCGCGCGCGCG 1632590

QY 1463 CGAACTAGTCAACCG 1522  
Db 1632591 CCAACAGCGCGCGGTTCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632650

QY 1523 TGCCATGGAACGCGCGCGCATACGAGATCGAGCTCGCATCGCTGGGACCGCGCGCA 1582  
Db 1632651 GCGTTGCGCGCGGTTCGCGCGGTGCGCGCGGTGCGCGCGCTGCGCGCGCGCGCG 1632710

QY 1583 ACGTCGCGCATCTCGTGGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1642  
Db 1632711 GCACCTCTGCAACCGCGCGGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632770

QY 1643 ACGGAGGAGACCTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1702  
Db 1632771 CGGCTACCGCGCGGTTCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632830

QY 1703 CCTACTCGGAGCG 1762  
Db 1632831 TTGGACCGCTTGGCGCGCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632890

QY 1763 TCGTCGACACCG 1822

Db 1632891 CCGCACCACCTGCTGCGCTCTGCTGGTGGTCTCAAGCGCTTACCGCTTGCCACCG 1632950  
QY 1823 AGGTCCACTTCGCGAGGCGACACGGGAATCTGCTCTACACCGAGCGCGCCCGCAC 1882  
Db 1632951 GCTCACCGCCACCGCGGCTCACCGTTGCGCGCTTGGCGCGTGGTGCCATCCGCG 1633010  
QY 1883 ACTTCCAGGCGATCGT 1898  
Db 1633011 CTGCCCCAGGCGCT 1633026

## RESULT 12

US-08-173-508-1

; Sequence 1, Application US/08173508

; Patent No. 5616485

; GENERAL INFORMATION:

; APPLICANT: Bartfeld, Daniel

; APPLICANT: Butler, Michael J.

; APPLICANT: Hadary, Dany

; APPLICANT: Jenish, David

; APPLICANT: Krieger, Timothy

; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED

; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND

; TITLE OF INVENTION: POLYPEPTIDES

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley &amp; Lardner

; STREET: 3000 K Street, N.W.

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/173.508

; FILING DATE: 23-DEC-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 18740/125/CACO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202 672 5300

; TELEFAX: 202 672 5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1908 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 146..1759

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 146..148

; OTHER INFORMATION: /note= "Met at position -39

; OTHER INFORMATION: represents fMet"

; FEATURE:

; NAME/KEY: sig\_peptide

; LOCATION: 146..262

; FEATURE:

; NAME/KEY: mat\_peptide

; LOCATION: 263..1756

; US-08-173-508-1

Query Match

Best Local Similarity 4.1%; Score 81.4; DB 1; Length 1908;

Matches 559; Conservative 0; Mismatches 721; Indels 15; Gaps 3;





Db 926 CACGTCGCGCGCATGGTCGTGACAGCGTGTCAACCCGTCGCCGCGACAAGATCTGGTAC 985  
Qy 1510 GGCAGCGCGCTCTCCATGGAACGCGCGCATATAGAGATCGAGCTCGACATCGCCCTGG 1569  
Db 986 CAGGCAACCTGGACAGGAGCTGCGCTTCGAGGCGCGCTGGAAGACTGGAGACTGG 1045  
Qy 1570 GACACGCGGAGAACGTGGCATCTCGGTGGCGCGCTCCCGGAGCGGAGCGGACACAG 1629  
Db 1046 GTGCGCGGAGAGCGCGCTTACACCTCGGCGGACACCGCGCGAGTTCAGGAGCAG 1105  
Qy 1630 AACATCGCAAGTACGAGAGCAGCTGTAGCTGCGAGCGGAGACCTCGACCTCGCGGG 1689  
Db 1106 TGGCTGAGCTGCG 1165  
Qy 1690 TACTCGCTGCGCGCGCTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1749  
Db 1166 GAGCTGATCTCTCTTCCAGAGCGCGCGCTACTAGGACTCGCGCTGGCGCGCGCGCG 1225  
Qy 1750 CACCTGCGCATCTGCTCGACACCGCAGAGCGTGTGAGGTCTCGTCAACGCGCGCGCACAC 1809  
Db 1226 GAGATCTTTCAGCAAGTACGTGCGCGCGGACACCGCGCGCGCTCGTGGAGCGCGCGCG 1285  
Qy 1810 GTGCTCTCCAGCAGTCTCCACTTTCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1869  
Db 1286 GACCTGTCGCGACCGCGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1345  
Qy 1870 GCGCGCGCGCGCACACTTCCACCGCGCATCTCGTCCG 1304  
Db 1346 GTGAGTGCACCG 1380

RESULT 14

US-08-387-942C-1  
; Sequence 1, Application US/08387942C  
; Patent No. 5939289  
; GENERAL INFORMATION:  
; APPLICANT: ERTESVAG, HELGA  
; APPLICANT: VALLA, SVEIN  
; APPLICANT: SKJAK-BRAEK, GUDMUND  
; APPLICANT: LARSEN, BJORN  
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES  
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/387,942C  
; FILING DATE: 09-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1809-106P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12588 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:  
; ORGANISM: Azotobacter vinelandii  
; STRAIN: E  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 290..1951  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2227..6438  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 6702..9695  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 9973..12588  
US-08-387-942C-1  
  
Query Match 4.0%; Score 80.4; DB 2: Length 12588;  
Best Local Similarity 42.3%; Pred. No. 1.7e-06;  
Matches 508; Conservative 0; Mismatches 691; Indels 3; Gaps 1:  
  
Qy 705 CACCGCGAACACGCGGAGGTTCGGCGCGCGCGGTCTCGCGCTCGCGACCCAGCCGAC 764  
Db 6875 CATCAAGAGCAACGTCCATATCTGTCGGCGCGGGGATGGCGAGAGCTCATCAAGCTGT 6934  
Qy 765 CGACGGCTTCGCGCAAGTACCAGAGCACTCTTACTTGTGTGACCGGCGGGTTCAC 824  
Db 6935 CGACGGCTGGGATCAGGACGTACCGGCGATCGTCGGCTCGGCTACGGCGAGGAGACCA 6994  
Qy 825 GTTACCGCCCTGCGCGACCCCGTCTGTCACACCGACCGGTGCGCGCGCCACCGCC 884  
Db 6995 CAACCTTCGCGATGAGCGACCTGACCTCGCGGCAACCGCGCAACACCGCGGCAAGGT 7054  
Qy 885 CGCGGAGATCGAGAACGCGGAGTGTTCGGCGACCCCAAGATCCACTGGGACACCGCCG 944  
Db 7055 CGACGGCTGTTCACGCGCTACATTCGCGCGGAGGACGCGCGCGGAGTGACCT 7114  
Qy 945 CG---GAGAAATGGGTCTCGGTCTCGGACGACTGCGGTGACGCGGCTTCTACACCTCG 1001  
Db 7115 GGAGCGGTGGAAATCCGTTGAAATGTCGGTTTACGTTTTCGATCCGCGACGAGCA 7174  
Qy 1002 GAACCTGCGGACGTGACACTTCGCGCAACTTCGACTACCGCAACCGACCGCTCGGG 1061  
Db 7175 CAACCTGACGATCCGCGACGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7234  
Qy 1062 CATCGAGTCCCCGACCTGTTTCGAGATCACCGGACGAGCGGAGACCGGCGGCTGGT 1121  
Db 7235 TTTCAGATCGCGGGGTGTTTCGAGAACACGCTCTGTAACAACGACCGCGCGCGGCTT 7294  
Qy 1122 CGCGCGGAGATGAGCGCTTACGGCATCGGCTTCCCATGAGCTAGCTTACTTGACAGG 1181  
Db 7295 CAACATGCTCACGACCAACGACTTCTGCTTGAGCAACACGCTCGGCTTTCGCAACGG 7354  
Qy 1182 CACTGGGACGCGGAGCAGTTCACGCGGACGACCTCACCGCGCAATGGCTCGACTGGG 1241  
Db 7355 CGGCGCGGCGCTGTTGTCAGCGCGGCTCGTCCGACGTGGCGGCGGCGGCGGCGG 7414  
Qy 1242 CTGGGACTGTACGCGCGCGCTCACCTGGCCATCGATCGACGCGCGCGGAGACCAACG 1301  
Db 7415 GATCGCGCGGCGCTTACTAGGACACGCGCTGGAAGCGGTGAGATCAAGATGCGCCA 7474  
Qy 1302 CGCATCGGTGGATGAACAACTGGAAGTACCGCGGACGCGGCGGCGGCGGCGGCGG 1361  
Db 7475 CGAGCTACCCCTCGAGAACGCGGAGATCTAGGGAACCGGCTATACGGGGTTCGCGGCT 7534  
Qy 1362 CGACGGCTACAAACGCGGAGAACTCGATCGTCCGCGAGCTGGGCTCGCGGCGGCGG 1421  
Db 7535 CGGCGCGGAGGATGTGCAGATCTCGCAACTATACCACTACCACTTCGAGAACGGTTC 7594  
Qy 1422 CGGTGTGTACACCTCTCTGTGAGCACCCCGCTGGCGAGCGCTGACGAACTACGTCAC 1481  
Db 7595 CTACGCGGAATCTCTCTGAGTCTCAGAGATACCGCGGCGGCTGTCGCGCAATTTCTA 7654







| Result No. | Score | Query % |        |    | Description       |                    |  |
|------------|-------|---------|--------|----|-------------------|--------------------|--|
|            |       | Match   | Length | DB | ID                |                    |  |
| 1          | 97.6  | 4.9     | 2712   | 10 | US-09-748-033-4   | Sequence 4, Appli  |  |
| 2          | 88.2  | 4.4     | 882    | 10 | US-09-974-300-735 | Sequence 735, Appl |  |
| 3          | 80.4  | 4.0     | 15872  | 9  | US-09-860-846-1   | Sequence 1, Appli  |  |
| 4          | 80.4  | 4.0     | 15872  | 9  | US-09-988-384B-1  | Sequence 1, Appli  |  |
| 5          | 80.4  | 4.0     | 15872  | 10 | US-09-861-289-1   | Sequence 1, Appli  |  |
| 6          | 80    | 4.0     | 1896   | 9  | US-10-124-880-15  | Sequence 15, Appli |  |
| 7          | 79.2  | 4.0     | 2334   | 10 | US-09-476-242-7   | Sequence 7, Appli  |  |
| 8          | 78.2  | 3.9     | 11220  | 9  | US-09-860-846-32  | Sequence 32, Appli |  |
| 9          | 78.2  | 3.9     | 11220  | 9  | US-09-988-384B-32 | Sequence 32, Appli |  |
| 10         | 78.2  | 3.9     | 11220  | 10 | US-09-861-289-32  | Sequence 32, Appli |  |
| 11         | 78.2  | 3.9     | 36778  | 9  | US-09-860-846-5   | Sequence 5, Appli  |  |
| 12         | 78.2  | 3.9     | 36778  | 10 | US-09-861-289-5   | Sequence 5, Appli  |  |
| 13         | 78.2  | 3.9     | 37948  | 9  | US-09-988-384B-3  | Sequence 3, Appli  |  |
| C 14       | 77.6  | 3.9     | 12441  | 9  | US-09-988-384B-3  | Sequence 3, Appli  |  |
| C 15       | 77.6  | 3.9     | 13613  | 9  | US-09-860-846-3   | Sequence 3, Appli  |  |
| C 16       | 77.6  | 3.9     | 13613  | 10 | US-09-861-289-3   | Sequence 3, Appli  |  |
| 17         | 75.8  | 3.8     | 2322   | 10 | US-09-476-242-18  | Sequence 18, Appli |  |
| 18         | 75.8  | 3.8     | 2322   | 10 | US-09-476-242-19  | Sequence 19, Appli |  |
| 19         | 75.6  | 3.8     | 4689   | 9  | US-09-860-846-34  | Sequence 34, Appli |  |

| Query Match           | 4.9%;           | Score 97.6;  | DB 10;     | Length 2712; |
|-----------------------|-----------------|--|------------|--------------|
| Best Local Similarity | 43.9%;          | Pred. No. 1.6e-13;   |            |              |
| Matches 650;          | Conservative 0; | Mismatches 794;  | Indels 36; | Gaps 4       |
| QY                    | 466             | CGGGGCTCGCTCCGTGCGGTCTTACACATGACGGCCCCCAGCGCTGGCTTCGCGACCCC    | 525        |              |
| Db                    | 845             | CGGCGCGCTGGCGCGCTCGTGGCGTCCGGTTCCTTCGCGGGATGGTCGGCCTC          | 904        |              |
| QY                    | 526             | CACGCGCCGGTACACACCCACGGCGCTTACAGCTGTACTTACCTGCACCTCGGACCCAGAAC | 585        |              |
| Db                    | 905             | GCGCCCCCACCGACGGCGCGCGAGCGCGCGACGCCCTCTTACACAGGAGCG            | 964        |              |
| QY                    | 586             | AACGGCCCCGGGGTGGGACACCGACGACCGACCGAGCGGTGCGCTTTCACGACCCAC      | 645        |              |
| Db                    | 965             | CAGGACTGGGCGACGGCTTCGAGGCGAAGTGGACGGTGAAGAACCGGCACCGCCCCC      | 1024       |              |
| QY                    | 646             | GGACCGTGATCGCGGTGGGGCCCGACTTCCCGGTGTGGTCCGGGTGCGCGGTGTCGGC     | 705        |              |
| Db                    | 1025            | CTCAGCGGCTGACCCCTCGAGTGGAGTTCGCCCGGGAACCAAGGTGACCTCGGCGTGG     | 1084       |              |

QY 706 ACCGGAACACGGCAGGTTTCGGCGCGCGGCTGCTGCGCTCGCGACCCAGCCGACC 765  
DB 1085 GACGCGACGCTACCAACACGCGACCACTGGACCGCAAGAACAGAGCTGGCGGG 1144  
QY 766 GACGGCTCGGCAAGTACCAGAGCAGTACCTCTACTGTGTCCACGACGGCGGTTTCAG 825  
DB 1145 AGCTTCGGCGCGCGGCTTCGGTTCAGCTTCGGCTTCACGGGACCGCGCGCGCCCGCC 1204  
QY 826 TTCACGGCGCTTCGGGACCGCTTCATCTGTCACACGCGGTTCGGCGCGCGCGCGCC 885  
DB 1205 TCGGGCTGCAAGCTCAACGGCGCTTCCTGCGAGCGGCGGCGTCCCGCGCACCGCG 1264  
QY 886 GCGGAGATCGAAGCGCGGAGTGTTCGCGCGACCCCAAGATCCACTGGGACACCGCGCGC 945  
DB 1265 CCCACGGCG 1324  
QY 946 GAGAGATGGTCTGCGTTCATCGGACGACTCGCGGTACGGCGCTTCTACACCTCGCGGAC 1005  
DB 1325 TGAAGCG 1384  
QY 1006 CTCGCGGACTGGACACTTCGCGCGCAACTTCGACTACCGGAAACCGCGCTTCGGCGCG 1062  
DB 1385 AAGGTGCGCACCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1444  
QY 1063 -----ATCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110  
DB 1445 TACTCTCTACTCGGTTCG 1504  
QY 1111 CACTGGGTGCTCG 1167  
DB 1505 GTCAAGGTACACCG 1564  
QY 1168 GCCTACTGGACGAGCG 1227  
DB 1565 AACCTCGGCTACTTCACGAGTGGGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCG 1624  
QY 1228 TGGCTGACTGGGCTGGGACTGTGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1287  
DB 1625 GTGACTCTCGGCTCG 1684  
QY 1288 GAGACAAAGCGCTCGCGATCGGTGATGAACACTGGAAGTACGCGCGCGCGCGCGCG 1347  
DB 1685 GCGGCAAGTGCACCATGGTGACAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1744  
QY 1348 CCCACGAGCATCGGCGGTACAAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1407  
DB 1745 GAGTCGGTTCGACGGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1804  
QY 1408 GCGGAGAGCGCTGGGCTGTGTACACCTCTGTGACACCGCGCGCGCGCGCGCGCGCG 1467  
DB 1805 CTCGCGCAAGCTCAAGCGCAAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1864  
QY 1468 TACGTACCG 1525  
DB 1865 ACCTGGTTCGGGCGGTTCACCGACCGCGGTGAAGACCGCGCGCGCGCGCGCGCGCG 1924  
QY 1526 CATGGAACG 1569  
DB 1925 CACGACCTGGTTCGAGGACCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1984  
QY 1570 GACACCG 1629  
DB 1985 GAGTACCG 2044  
QY 1630 AACATCGGCAAGTACGAGGACGAGCTGTACGTTCGACCGCGCGCGCGCGCGCGCGCG 1689  
DB 2045 ATGGTTCGAGCG 2104  
QY 1690 TACTCGCTCG 1749  
DB 2105 GACGCGAGCTCGCGGCGCAAGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2164  
QY 1750 CACCTGCGCATCTCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1809

DB 2165 GACTGTACAACGTGATGACGTACGACTTCTTCGGCGCGCTGGGACAAGACCGCGCGAC 2224  
QY 1810 GTGCTCTCCACGACGCTCCACTTCGCGAGGCGCGACACGGGAATCTCGCTCTACACCGAC 1869  
DB 2225 GCG 2284  
QY 1870 GCG 1909  
DB 2285 GCGCGCATCGCCCAAGCTCAAGGCGAAGGCGGTCCCGCGGA 2324

## RESULT 2

US-09-974-300-735  
; Sequence 735, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085-500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 735  
; LENGTH: 882  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-735

Query Match 4.4%; Score 88.2; DB 10; Length 882;

Best Local Similarity 63.4%; Pred. No. 2.4e-11;

Matches 135; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1055 TCGCGCGCATCGAGTGCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1114  
DB 654 TCGGCATCATCGAATGTCGGGATCTGTATAGATCGGGCGGATGACGGAACATTAAT 713  
QY 1115 GGTGCTCG 1174  
DB 714 GGGTCTTGGAGCAAGCGCCAAACGGAAGGACCGGCAAGCGGAATCTTATGCACT 773  
QY 1175 GGACAGGCG 1234  
DB 774 GGACAGGCG 833  
QY 1235 ACTGGGCTGGGACTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1267  
DB 834 ATTACGGATTGACTGCTGATGCGGACGTGACAT 866

## RESULT 3

US-09-860-846-1  
; Sequence 1, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860,846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 15872

TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-860-846-1

Query Match 4.0%; Score 80.4; DB 9; Length 15872;  
Best Local Similarity 42.7%; Pred. No. 1e-09;  
Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps

|    |      |  |      |
|----|------|--|------|
| Db | 1372 | CGTCGTCACCGCGCGCGGCGAGGCCCTGC- - - - GCGCCCAAGCGCCGCGGCGCTGCACGA | 1428 |
| Qy | 1446 | CCCGGTGGCAGCGCTGACGAACCTAGCTACCGCCACCAACACACACTCCCGAGACCGAACGCT  | 1505 |
| Db | 1429 | GGCGGTCGAAGCGGACCCGAGCTCGCGCGCCGCGCACTCGCCCGTGCCTGGTCACCCAC      | 1488 |
| Qy | 1506 | CGACGGCAGCGCGTCTGCCATGGAAACGGACCGCATACGAGATCGAGCTCGACATCGC       | 1565 |
| Db | 1489 | CCGTACGGTCTTACGCAACCGGTGCGTCTCTCGCCCGGACCGCGCCGCTCTCTGGA         | 1548 |
| Qy | 1566 | CTGGGACACCGCGACGAACGTGGCATCTCGGTGGCGCGCTCCGCCGACGAACCCGGCA       | 1625 |
| Db | 1549 | CGGCTTCGGCGCCCTCGCGCGCGGACCGCGCGCGGTGTTACACGGCACCCCGCG           | 1608 |
| Qy | 1626 | CACGAACATCGGAAGTACGGAGCAGACCTGTACTGACGAGGACGCTCCGACCTCGC         | 1685 |
| Db | 1609 | CCCGGGCGCTCGCGCTCTGTTCAGCGCGCCAGGGTGCCCAACCTACGGGCATGGGCAT       | 1668 |
| Qy | 1686 | CGGCTACTCCTCGCCCGCTACTCGGAGCGCGCGCGCCCATCGACCCCGCGGCCGATC        | 1745 |
| Db | 1669 | GGAGTTGTAGCGCGCGCCACCCCGCTTCGCGAGCGCTTCGACGCGCTCGCGCGCGAAT       | 1728 |
| Qy | 1746 | CGTGCACCTCGCATCTCGTCACACACCGAGCGCTCGAGGTTCTGTCACCGCGGCCA         | 1805 |
| Db | 1729 | GGACCCCTCTCGACGGGCCCTCGCCGAATCTGTGCGGGCGGACACCTCGACCG            | 1788 |
| Qy | 1806 | CACGTCGCTCTCCACGAGGTCCACTTCGCGGAGGCGACACGGGAATCTCGTC- - - -      | 1860 |
| Db | 1789 | CACGTCACACACAGCCGCGCTCTTCGCCGTGGAGTTCGCGCTCCACCGCTCGTCGA         | 1848 |
| Qy | 1861 | -TACACGAGCGGGCGCCGCACATTCACCGGCATCTGTCGCGGAGATTCGCCAGCG          | 1919 |
| Db | 1849 | GTCTGGGGGTACGCGCGCACTGTCTCGCGGCCCATCTCGTCGCGAGATCAGCGCGC         | 1908 |
| Qy | 1920 | GATCTAGCGCATGCACACACACCGCTCACCGAACCGCGCCCGGGAGACGACGGCGGA        | 1979 |
| Db | 1909 | CCAGTCGCGGGGTCTGTGCTGGCGACGCCCGCCGCTCTGTCGCGCGCGCGCGCG           | 1968 |
| Qy | 1980 | CAATGCACGTCCTCGTCG   | 1999 |
| Db | 1969 | CCTCATGACGGCGCTCCCG  | 1988 |

## RESULT 4

US-09-988-384B-1 ; Sequence 1, Application US/09988384B ; Publication No. US20030073824A1

APPLICANT: Sherman, D. H.

APPLICANT: Lin. H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

**TITLE OF INVENTION:** DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.536US1

; CURRENT APPLICATION NUMBER: US/09/988,384B

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: PCT/US99/14398

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 53

; SEQ ID NO 1

; LENGTH: 15872

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-988-384B-1

Query Match 4.08: Score 80.4: DB 9: Length 15872:

Best Local Similarity 42.78; Pred. No. 1e-09

| Matches | 649 | Conservative | 0 | Mismatches | 856 | Indels | 15 | Gaps |
|---------|-----|--------------|---|------------|-----|--------|----|------|
| Matches | 649 | Conservative | 0 | Mismatches | 856 | Indels | 15 | Gaps |

QY 492 CATGACGCCGCCCCAGCGGCTGGCTCTGGACCCCCAACGCCGGTCAACCAACCCAGGCGC 551

Db 472 CATCGCCAAACCCGGTCTGTACCACTCTGGCCCTGAGGGCCGAGCCCTACCGTGCAGCC 531  
QY 552 CTACAGCTGTACTACTCGACTCGACCAAGAACAGCCGCCGGGTGGGACAGC 611  
Db 532 CGCGAGTGTCTCTCTGCTGCGCGTGCACCTGGCCCTGCGAGTCCCTGCGCGCGGGA 591  
QY 612 GAGCAGGACCGGCTGCTGCTTACAGGACACAGGACCGGCTGAGTGCCTGCGGCCGA 671  
Db 592 GTCCAGACGGGCTGCTGCGCGGGTGAACCTCAACATCTTCGCGGAGAGCGCGTAC 651  
QY 672 CTTCCTCCGTGCTCGGCTGCGGCTGCTGCGGACCGGACAGCGAGGTTTCGGCG 731  
Db 652 GAGAGAGCGCTTCGGTGGACTCTCCCGAGCGGACCGCTTACCTTCGAGCGCGGCG 711  
QY 732 CGCGCGGCTGCTGCGGCTGCGGACCGGACCGGACCGGCTGCGGACCGGCTGCGGACG 791  
Db 712 CAACGATTCTCGGCGGAGGCGGCGGAGTCTGCTGCTCAAGCGCTCTCCCGCGC 771  
QY 792 GTACCTCTACTGCTGACGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 851  
Db 772 CTTCCCGGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831  
QY 852 CTTCAACACCGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 908  
Db 832 AGCCACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 891  
QY 909 GTTCGCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 968  
Db 892 GTACCGGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 951  
QY 969 AGGACTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028  
Db 952 AACCCCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1011  
QY 1029 CAATCTGACTACCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1088  
Db 1012 CCCCAGGAGGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1071  
QY 1089 CACCGCAG---ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1145  
Db 1072 CCGCGCGGATGCTGCGGCTCATCAAGAGGCTCTGCGGCTGCGGCGGCGGATCCC 1131  
QY 1146 CACCGGCTCCCATGACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1205  
Db 1132 GCGAGGCTCAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191  
QY 1206 CCGCGGACGCTACCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1265  
Db 1192 CTTGCGGAGGCTGCGGAGTGGCGGACCGCGGAGCGGAGCTCTGCGGCGTCTAG 1251  
QY 1266 CTTGCGGATGCTGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1325  
Db 1252 CTTGCTGCGGATGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1311  
QY 1326 GAGTACCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1385  
Db 1312 CCGGAGGAGCGGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1371  
QY 1386 GATCTCTCGGAGTCTGCGGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1445  
Db 1372 CTTGCTGAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1428  
QY 1446 CCGCTGCGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1505  
Db 1429 GCGCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1488  
QY 1506 CGAGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1565  
Db 1489 CCGTACGCTTTCAGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1548  
QY 1566 CTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1625  
Db 1549 CCGCTCTGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1608

QY 1626 CACGAACATCGCAAGTACGGAGGAGAGCTGTACGTGACCGAGGAGCCCTCGAGCTCGC 1685  
Db 1609 CCGCGGCGGCTCGCGCTCTGTTCAGCGGCGCAGGCTGCCAACGTACGGGATGGGAT 1668  
QY 1686 CCGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1745  
Db 1669 GAGTGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1728  
QY 1746 CCGTACCTGCGCATCTCTGCTGACACCGGAGGCTGAGGTCTTCTGTCACGCGCGG 1805  
Db 1729 GAGCCCTCTCTGACGCGGCGGCTGCGGAGCTGCTGCGGCGGCGGAGACCTCGAC 1788  
QY 1806 CACCGTGTCTCTGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860  
Db 1789 CACCGTCTACACAGCGCGGCTCTGCGGCTGAGGCTGCGGCTGCGGCTGCTGCTG 1848  
QY 1861 TACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1919  
Db 1849 GTCTGCGGCGGCTGACGCGGAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 1908  
QY 1920 GATCTAGGCTGATGACACACCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 1979  
Db 1909 CCAGTCTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1968  
QY 1980 CAATCGACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1999  
Db 1969 CCTCATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1988

RESULT 5  
US-09-861-289-1  
; Sequence 1, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-1

Query Match 4.0%; Score 80.4; DB 10; Length 15872;  
Best Local Similarity 42.7%; Pred. No. 1e-09; Indels 15; Gaps 4;  
Matches 649; Conservative 0; Mismatches 856;

QY 492 CATGACGCGCGCGCGGCTGCTGCGGAGCCCGGCGGCTGCGGAGGCGGCGGCGGCGG 551  
Db 472 CATGCGCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531  
QY 552 CTACAGCTGTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 611  
Db 532 CCGGAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591  
QY 612 GAGCAGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 671  
Db 592 GTCCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651  
QY 672 CTTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731  
Db 652 GAGGAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 711

QY 732 CGGCGCGGTCTCGCGGTTCGGAGCCAGCCAGCGAGCGGTTCGGAGTACAGGAGCA 791  
Db 712 CAACGGATTCGTCCGGGCGAGGCGCGGAGTCTGTACTAAGCGGTCTCCCGCGC 771  
QY 792 GTACCTCTACTGTGTCAGCCAGCGCGGTTCACGTTTACCGCCCTCCCGACCCCGTCAT 851  
Db 772 CCTCGCGAGCGGACCGTGTCCAGCGGTCTATCCGCGCCAGCGCGGTCAACAACGACGG 831  
QY 852 CGTCAACACCGACAGGTTCGGCGCCAGCCAGCCCGCC---GAGATCGAGAACGCCGAGTG 908  
Db 832 AGCCACCCCGGTCTACCGGTTCGCGAGCGCGCCAGGAGAGGTGTGTCGCGGAGGC 891  
QY 909 GTTCGGGACCCCAAGATTCACCTGGGACACCGCCCGCGGAGAAATGGGTCTGCGTCATCGG 968  
Db 892 GTACCGGAAGCGCGCTTGGACCGGTTCGCGGTCCAGTACGTCTCAAGTCCACGGCACCGC 951  
QY 969 ACGACTGGGTACCGCGGTTCACACTCGCCGAACCTGCGCGACTGGACACTTCGCGC 1028  
Db 952 AACCCCGGTGCGGACCCATCGAGGCGCGCGCTCGGCGCCGTCTCGGCTCGGGCGC 1011  
QY 1029 CAACTTCGACTACCCGGAACACCGCCCTCGCGGCGATCGAGTGCCTCCGACCTGTTTCGAGAT 1088  
Db 1012 CCCGCGGACGAAACCCCTGCTGCTGCGTCCGCCAAGACGATCGCGGACCTCGAAG 1071  
QY 1089 CACCGAG---ACGACGGGACACGCCACTGGGTGCTCGCGCCAGCATGAGACGCTACGG 1145  
Db 1072 CGCGCGGATCGTCTCGGCTCATCAAGACCTCTCGCGCTCGCGCGCGCGCGATCCC 1131  
QY 1146 CATCGGCTCCCATGACGTACGCTACTGGACAGGACCTGGGACCGCGGAGCATTTCCA 1205  
Db 1132 GCGAGGCTCAACTTCGCTAGCGCCCAACCGGACATCCGCTCGACACCTCGGGCTCGA 1191  
QY 1206 CGCGGAGGACTCAACCCGCAATGGCTCGACTGGGGTGGGACTGGTACGCGCGCGTCAC 1265  
Db 1192 GGTCCCGACGGCTGCGGGAGTGGCGCACCGGACCGCAACTCTCGCGCGGTGAG 1251  
QY 1266 CTGGCCATGATCAACGCGCGGAGACAGAGCGCTCGCCATCGCGTGGATGAACAATG 1325  
Db 1252 CTCGTTGCGGATGGCGGACCAACGCGCCAGCTGCTCTAGCGAAGCGCCCGCCAGG 1311  
QY 1326 GAAGTACGCGCAGCGACGTCCCGACGAGCATCGCGGCTACACGGGAGCAACTC 1385  
Db 1312 CGCGGAGCGCGCGGATCGATGAGGAGACCCCGCTCGACAGCGGGCGCGACTGCGCTT 1371  
QY 1386 GATCGTCGCGAGTTCGCGGTTCGCGCGACAGCGCTGGGTGATACACCTCTCTGAGCAC 1445  
Db 1372 CGTGTCAACGCGCGCGGAGGCGCTTCG---GGCGCCAGCGCGCGCTGACGA 1428  
QY 1446 CCGCGTGGCAGCGGTGACGAATACGTACCGGACACCACTATCCCGACCGGACCGT 1505  
Db 1429 GCGCGTCGAAGCGGACCGGAGCTCGCGCGCGCGGCACTCGCGCGGTGCTGTCACAC 1488  
QY 1506 CGAGCGGACGCGCTCTGCTGATGGAACGAGCGGCGATACGATCGAGTCTGACATCGC 1565  
Db 1489 CCGTACGCTTTCACGACCGGCTGCTGCTCGCGCGGACCGCGCGCGCTCTCGA 1548  
QY 1566 CTGGGACACCGGACGAGCTCGGCTATCTGGTGGCGGTTCCTCCCGACGAAACCGGCA 1625  
Db 1549 CGGCTCGGCGCGCTCGCGCGCGGAGCGCGCGCGCGGTGTACCGGACCGCGCGC 1608  
QY 1626 CACGAATCGGCAAGTACGAGGACGACACTGTACGTGACCGGAGGACCGCTCCGACCTCGC 1685  
Db 1609 CCGCGGCGCTCGCGCTCTGTTTACGCGGCGAGGTTGCCAAGCTACGGGATGGGAT 1668  
QY 1686 CGGTACTCTCGCCCGCTACTCGGAGCGCGCGCGCGCGCTGACCGCGCGCGCGCGATC 1745  
Db 1669 GGAGTTGACGCGCGCGCGCGCTTTCGCGAGCGCGCTTTCGACCGCGCTCGCGCGCAACT 1728  
QY 1746 CGTGCACCTGCGATCTCTGTCGACACCCAGAGCGGTCTGAGTCTTCGTCAACGCGCGGCA 1805  
Db 1729 GGACCCCGCTCTGACCGCGCGCGCTCGCGGAGTCTGTCGCGGCGGCGGACACCTCGACCG 1788  
QY 1806 CACCGTCTCTCCAGCAGGTTCACCTTCGCGGAGGCGGACACGGGAATCTCGCTC----- 1860

Db 1789 CACGCTCCACACACAGCCCGCGCTCTTCGCGGTGGAGGTGCGCTCCACCGCTTCGTCGA 1848  
QY 1861 -TACACGAGCGCGCGCGCACACTTCACCGCATCGTCTGCGGAGATTTGGCCAGGC 1919  
Db 1849 GTCTGCGGCGTACAGCGGACCTGTCGCGGCGCACTCCGTCGCGGAGATTCAGCGCGC 1908  
QY 1920 GATCTAGCGATGACACACACCTGCTACCGGAAGCGCGCGCGGAGAGACGAGCGCGA 1979  
Db 1909 CCAGTCCGCGGCTCTGTCGTCGCGAGCGCGCGCGCTCTGTCGCGCGCGCGCGC 1968  
QY 1980 CAATCGACAGTCTCTCTCG 1999  
Db 1969 CCTCATGACGCGCTCCCG 1988

RESULT 6  
US-10-124-880-15  
; Sequence 15, Application US/10124880  
; Publication No. US20030026810A1  
; GENERAL INFORMATION:  
; APPLICANT: Jorgensen, Per Lina  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Andersen, Lene No. US20030026810A1boe  
; APPLICANT: Schulein, Martin  
; APPLICANT: Outtrup, Helle  
; TITLE OF INVENTION: No. US20030026810A1el Rhamnogalacturonan Hydrolases  
; FILE REFERENCE: 5572.204-US  
; CURRENT APPLICATION NUMBER: US/10/124.880  
; CURRENT FILING DATE: 2002-04-19  
; PRIOR APPLICATION NUMBER: US/09/311.626B  
; PRIOR FILING DATE: 1999-05-13  
; PRIOR APPLICATION NUMBER: 0608/98  
; PRIOR FILING DATE: 1998-05-01  
; PRIOR APPLICATION NUMBER: 60/084.358  
; PRIOR FILING DATE: 1998-05-05  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 1896  
; TYPE: DNA  
; ORGANISM: Streptomyces coelicolor  
US-10-124-880-15

Query Match 4.0%; Score 80; DB 9; Length 1896;  
Best Local Similarity 43.8%; Pred. No. 1.6e-09;  
Matches 665; Conservative 0; Mismatches 825; Indels 30; Gaps 6;

...QY 489 CCACATGACGCGCGCGCGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548  
Db 276 CAACCTCCAGCGCGCTCACCGGCTCCACACCTACTTCCACTCCGCGCGCGCGCGCGCGCG 335  
QY 549 GCGCTACAGCTGTACTACCTGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 608  
Db 336 CGACTACAGCTGCTGCGCGCGCTGCTGGAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 395  
QY 609 CGCGAGCACGACGACGCGCGCTGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 668  
Db 396 CCACTTCGCGCGCGCTGCTGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455  
QY 669 CGACTTCGCGCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728  
Db 456 CGAGCGGCTCTCTTACACTTACGAGGCGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCG 515  
QY 729 CGCGCGCGCGGTGCTGCTGCG 788  
Db 516 CGCGCGCTCTGACCTGCTGCTCAAGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575  
QY 789 GCAGTACTCTACTGCTGCGAGCGCGGTTCACGTTTACCGCTTACCGCGCGCGCGCGCGCG 848  
Db 576 CGGCTACACCGGCAACAGGTGCTGCGAGCGGCTCAAGTCAAGTCAAGTCAAGTCAAGTCA 632  
QY 849 CATGCTCAACACCG 908

Db 633 GCGGTGACCTGGCGCGCAACATCCGCTCCGGCGCCACATACACCCAGTTCCAGGTGA 692  
QY 909 GTTCGGCGACCCAAAGATCCACTGGGACACCGCCCGCGAGAATGGTCTCGGTTCATCGG 968  
Db 693 CGACTACGACGCGAGCGCGCGCGAGGTGCCATAGACCGCGCGACCGCAAGA 752  
QY 969 AGCACTCGGGTTACGCGGGTTTACACCTCGCCGAACCTGGCGGACTGGACATTCGCGG 1028  
Db 753 CGGCACCGCGCGGTTCATCGGCAACTCTCGCGGATCACCACCTCGAGCGGTACGT 812  
QY 1029 CAACCTGACCTACCCGACACCGCCCTCGGGGATCAGTGGCCCGGACCTGTTCCAGAT 1088  
Db 813 CCTCCCGCGCGCAATACCTACCATGTTCAACGCGCGGACCGCACCGCATGGGAC 872  
QY 1089 CACCGACGACGCGGACACGCACTGGTGTCTGCC---GCCAGCATGGAGCGCTACGG 1145  
Db 873 CGTCGACTACCTCCGCGCGCGCGGTCTCGTCTCTGGGCGACTCTTAGGGCAACG 932  
QY 1146 CATCGGCTCCCATAGCTAGCTACTTGACAGCACTGGAGCACTGGGAGCGGACAGTTCCA 1205  
Db 933 COTGACCGGTTCTTGGGGGACGCGGTACCTGGACGCTCCCGCCCTCCGTGATTAT 992  
QY 1206 CGCGACGACCTACCCGCAATGCTCGACTGGGCTGGGACTGTAGCGCGCTCAC 1265  
Db 993 GCGCGCGGTACTACCGCGACCGGTGATCGCGCTTGGGACTGG---CGGGACGGCG 1049  
QY 1266 CTGGGCATCATGATCGAGCGCGCGAGACCAAGCGCTCGCCATCGCGTGGATGAACAATG 1325  
Db 1050 GTTACCGCGCGCTGACTTTCGACACCAACTCTCCACCAACACGCGCAAGGCTACGA 1109  
QY 1326 GAAGTACGCGGACCGGAGTCCCGACCGCATCGAGCGGTACAGCGGCGAGAACTC 1385  
Db 1110 CGGCGAGGCAACCAACAGCTCTCGCTCGGAGACGTGACGGTGTAGCGCGGAGAGAT 1169  
QY 1386 GATGCTCGCGAGTGGCTGCGCGACAGCTGCGGCTGGTACACCTCCTGAGAC 1445  
Db 1170 GGTCTACGCGCGGATGGCGGTGAGAGACACAGGCTTACGCGCTTGGACACCAAGAA 1229  
QY 1446 CCGCGTGGCAGCGTACGAACTAGCTACCGGCCACCAACC-----ACACTCCCGGACG 1499  
Db 1230 CCACGCGACGCCATGACGTGCGGCACTCGACCCGTCGCGGCGGCTCGGAGGATT 1289  
QY 1500 GACCTGACGCGCAGCGCTCTCCATGAGACGCGCATACGAGATGAGACTGA 1559  
Db 1290 CAAGTCTGACGAGGAGCGCTCGAAGCCCTCGTGTACCTGGCGGACCGCCGACGCGCA 1349  
QY 1560 CATCGCTGGGACCGCGACGAGCTCGGCATCTCGGTGGGCGCTCCCGGACGGAAC 1619  
Db 1350 GATCTCTGTCCACCGGCGGAGCGCGACACGCGCGCGGTGTCTCGGGGACATCTG 1409  
QY 1620 CCGGCACACGAATCGGCAAGTACGAGCAGACCTGTACGTGACCGAGGACCTCCGA 1679  
Db 1410 GTCCGCGACGCGCGCGCGAGTCTGTGTC-----GTCCGCGGAGAGCGCATCCGAA 1463  
QY 1680 CCTCCCGGTTACTCGCTCGCCCTACTCGGAGCGCGCGCCCATCGACCCGCGCG 1739  
Db 1464 CCCCAGGGGACCGCTGCTGCGGAGCGGACCGACCCCTCCAGCGCAACTTCTTTCTGTG 1523  
QY 1740 CCGATCCGTGCACTGCGCATCTCTGTCGACACCGACAGAGCTCGAGGTCTTCGCAACGC 1799  
Db 1524 GAGGGGACACCTGCTGTAAGTCTCTGCGGACACCGACCCAGTGCAGACAGTACGGCACCTC 1583  
QY 1800 CGGCGACACCGTGTCTCCAGCAGTCCACTTCGCGAGGCGGACACGGAATCTCGCT 1859  
Db 1584 GGGCGACA-----CCCGCTCTACCGGCTCCGCGCTCGGCTCCACACAGGCGAC 1634  
QY 1860 CTACACGACGCGCGCGCGGACACTTACCGGCACTCTGCTCGCGGAGATTGGCCAGCG 1919  
Db 1635 CAAGGCCACCGGCTCTTGGCGGCGAGATCTCTCGGACTGGCGGAGAGGTCTGTG 1694  
QY 1920 GATCTAGCGGATGACACCAACCGCTCACCGAAGCGCGCGCGCGGAGACGCGCGCA 1979

Db 1695 GCGACGTCGACACACACGCGCCCTGCGCATCTACTCCACCCCTACGACACGACACCG 1754  
QY 1980 CAATCGACACGTCCTCTGTCG 1999  
Db 1755 CATCAGACGCTCTCTCCAG 1774  
RESULT 7  
US-09-476-242-7  
; Sequence 7, Application US/09476242  
; Patent No. US20020146683A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: HARTOG, Karin  
; APPLICANT: MARTIN, Eric  
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
; FILE REFERENCE: 1605.002  
; CURRENT APPLICATION NUMBER: US/09/476.242  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 2334  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Leu122-Ser199  
US-09-476-242-7

Query Match 4.0%; Score 79.2; DB 10; Length 2334;  
Best Local Similarity 43.8%; Pred. No. 2.4e-09;  
Matches 458; Conservative 0; Mismatches 573; Indels 15; Gaps 2;

QY 795 COTCTACTGTGTCGACGCGGGTTCACGTTTACCGCCCTGCGCGGACCGCTCATCGT 854  
Db 147 COTGTTCTGCGCCAGCGACGCAAGCCTACGACACCGAGGTGCACAAGCTGTGGCCAC 206  
QY 855 CAACACGACGGTTCGCGCGCCACCGCCCGCGAGATCGAGACGCGGAGTGTTCG 914  
Db 207 CCACGCTTGTGCGCCAGCGACCGCCCGAGGATCGTCTGGAGACGTGACCGA 266  
QY 915 CGACCCCAAGTCCACTGCGACCGCGCCCGGAGAAATGGTCTCGTCTACGGAGCT 974  
Db 267 GAATTCACATGTGGAAGAACATGTTGGAGCAGATGACGAGGACATCATCAGCT 326  
QY 975 GCGGTACCGCGCTTCTACACTCGCCGAACTGCGGACTGGACCTACGGCGCAACTT 1034  
Db 327 GTGGACACAGCGCTGAAGCCTGCTGAAGCTGGGCAACAGCGTGTACCCAGGCGCTG 386  
QY 1035 CGACTACCGGACCGCGCTCGGCGGCGATCGAGTGGCCCGGACCTGTTGAGATCACCG 1094  
Db 387 CCCCAGGTGAGTTCGAGCGCCATCCCATCTACTGCGCGCCCGCGCGCTTCGCCA- 445  
QY 1095 AGACGACGGGACACCGCCACTGGGTGCTCGCGCCAGCATGGACCTACGGCATCGGCT 1154  
Db 446 -----TCTGTAAGTCAACAGACAGAAAGTTCAACGCGAGCGCCCTGCAACAGCTGAG 500  
QY 1155 CCCCATGACGTACGCTTACTGACAGGACCTTGGACGCGGAGAGTTCACGCGCGACGA 1214  
Db 501 CACCGTCAGTGCACCCACCGCATCCGCCCGGTGTGAGCACCCACGCTGCTCTCAACGG 560  
QY 1215 CTTACCCGCAATGGTCTGACTGGGCTGGGACTGGTACGCGGCGCTACCTTGCCCATC 1274  
Db 561 CAGCCTGCGCGAGGAGCGGTGGTATCCGACGAGAGAACTTACCGACAGCGCTAAGAC 620  
QY 1275 GATCGACGCGCGGACGACCAAGCGCTCGCATCGGTGGATGAACAACCTGGAAGTACGC 1334  
Db 621 CATCATCTGAGCTGAAGGAGAGCGTGGAGTCACTGACCCGCCCAACACAAACAC 680  
QY 1335 CGCAGCGGACGTCCCGACCGGATCTCGAGGCTTACAGCGGCGAGAACTCGATCTCGG 1394  
Db 681 CCGAAGAGCATCACCATCGCGCGCGCGCTTCTACGCCACCGGCGGACATCATCGG 740





Db 4000 GTGACCTCGGTGGGAGCCCTGGGAGGCGACGCCGGTCAACGAGGGTGGGAGCCGGG 4059  
QY 1780 GTGAGGCTTTCGTAACCGCGCGCCACACCGGTCTCCAGCAGGTCCACTTCGCGGAG 1839  
Db 4060 GAGCGGCTGCGCGCTCGGCTGCGCCCTCGCCCGCGGACGCGCTCAACCGCCCTG 4119  
QY 1840 GCGGACCGGGAATTCGGTCTACAGCGGCGCGCGCGGACACTTCACCGG 1892  
Db 4120 GACACCGGCTCGGCGACGCGGACACCGCGCTCAGCATCGCGGAGTGCAGTGS 4172

## RESULT 9

US-09-988-384B-32

; Sequence 32, Application US/09988384B

; Publication No. US20030073824A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.536US1

; CURRENT APPLICATION NUMBER: US/09/988.384B

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: PCI/US99/14398

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 53

; SEQ ID NO 32

; LENGTH: 11220

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-988-384B-32

Query Match 3.9%; Score 78.2; DB 9; Length 11220;  
Best Local Similarity 41.9%; Pred. No. 3.4e-09;  
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCTCGCTCGGTGCGCTTACACATGACGCCGCCCGCGGTGCTCTCGACACCCCAA 528  
Db 2740 GACTCCCCCGGTCAGGGTCCGTGAGGACTCTTGGCGGTACCGACTGGAAG 2799  
QY 529 CGCCGGTACACACCGCGGCGCTTACAGTGTACTACTGCTACCTCGACAGAACAC 588  
Db 2800 CTTCTCGGGTGGCGAGCGCTCCAGCGCGCGGCTGTCGGGCGTGTGCTCGTC 2859  
QY 589 GSCCGCGGCGGTGGACACGAGCAGCAGCAGCGGCGTCCCTTACGACACACGCG 648  
Db 2860 GTCCCGGAGGACCGTTCGCGGAGCGCCCGCGGTGCTGCGCGCGCTGTCCGGCGCG 2919  
QY 649 ACCGTGATCGCGCTCGCGCGCGACTTCCCGGTGCTGCGGTGCGGCGGTGCTCGGC 708  
Db 2920 GCGACCCGCTACACTGAGCTGTGCTGCGGTGCGGCGGCGGCGGCTGCGCGCGAG 2979  
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGCGGCTGCTGCGGCTCGACACCGCGGCG 765  
Db 2980 GTGGCGGAGGCGCTGCGCGCGCGCGGTGGAGCGCTCGAGCGGCTCTCTCGTGTGCG 3039  
QY 766 GACGCGCTCCGCAAGTACAGGAGCAGTACCTTACTTGTGACCGCGGCGGTTCAGG 825  
Db 3040 TGGGACGAGAGCGCGACCCCGCGGACACCCCGCGCTTACCCCGGCGGCGGCGCG 3099  
QY 826 TTACCGCGCTCGCGGACCGCGCTATCGTAACACCGCGGTGCGCGCGCGCGCGCGCG 885  
Db 3100 CTCACCTGCTGCGCGCTGAGGACCGCGCGGTGCGCGCGCGCGCTGTGGTGTGCG 3159  
QY 886 GCGGAGATCGAGACCGCGGAGTGTTCGCGGACCGCGCGGCTGCGGACCGCGCGCG 945  
Db 3160 CACGCGCGGCTGCTGCGCGCGCGCGGTGCGGACCGCGCGGCTGCGCGCGCGCGCG 3219  
QY 946 GGAGAATGGGTCTGCTCATCGGAGCTGCGGTACGCGCGGCTTCTACACCTGCCGAG 1005

Db 3220 TGGGCGATGGGCGGCTGCGCCCTCGGAGCACCCCGAGCGGTGGGCGGCGCTGTATCGAC 3279  
QY 1006 CTGCGGAGCTGACACTTTCGCGGCAACTTCGACTACCGGAACACCGCCCTCGCGCGCATC 1065  
Db 3280 CTGCGCTCGGAGCGCGGCGGCGGCTTGACCGCATGACCAACGCTCTCTCCCGCGGT 3339  
QY 1066 GAGTCCCGGAGCTTTCGAGATCACCGAGACGAGGAGGACAGGCACTGGGTGCTCGGC 1125  
Db 3340 ACGGCTGAGGACGAGTGGGTAGCGGCTCTCGGCGCTCTCGCGCGCGCTCTCGTCGC 3399  
QY 1126 GCGACATGGAGCGCTACGGCATCGGCTTCCCATGAGTACGCTACTTGGACAGGACG 1185  
Db 3400 GCTTCTCCCGCGCACGCGACGCTTTCGCGGTGGCAGCGCGACGCGCATGTGCTC 3459  
QY 1186 TGGGACGCGGAGCTTTCACCGCGGAGCTTTCACCGCGGAGGCGGCTGCGACTGGGCTGG 1245  
Db 3460 GTCACCGGTGCGGAGGAGCTTTCGCGCGGCGGAGGCGGCGGCTGCGCGCGGAGCG 3519  
QY 1246 GACTGTAGGCGGCGCTACCTTGGCCATGATGAGGCGCGGAGACCAAGGCGCTCGGC 1305  
Db 3520 GCGGAGACACTCTCTCTCCACACACCCCTCTCGGAGCGGAGGCGCGGAGGCGACCTCC 3579  
QY 1306 ATCGCGTGGATGAACAACCTGGAAGTACCGCGCAGCGCTCCCGACGACCATCCGAC 1365  
Db 3580 GGTGCGCGGAGGACTTCCGCGCTCCCGGCTGCTGCGGAACTCGCGGACTGGGCGG 3639  
QY 1366 GGCTACAAACGGGCGAGAACTCGATCTCGCGGAGCTGCGGCTCGCGCGAGC-----CT 1419  
Db 3640 ACGGCGACCTGCTGACCTTGGACCTACGAGCGGAGGCGGCGCGCGCTGCTCGGC 3699  
QY 1420 GCGCGCTGTACACCTTCTGAGCACCCCGGTGGAGGCGCTGAGAACTACCTACCGCG 1479  
Db 3700 GCGCTCTCGGCGCGACCGCTCAGCGCGCTCTCCACCTCGCGCGCGGCTGCGACTCC 3759  
QY 1480 ACCACACACTCCCGACGCGGAGCTGACGCGCGGCTCTGCGATGGAAACGACG 1539  
Db 3760 GAGCGCTGCGCGCGGAGCGGCTGCGCGGCTGCTGCGCGGAGGCGCGAC 3819  
QY 1540 GCATACGAGATGAGCTGCGATCTGCGGACACCGGAGCGGAGCGGCTGCGGCTGCGGTG 1599  
Db 3820 GCGCGCTCCACCTGCGGCGGCTCTCGGAGGCGGCGGCTGCGGAGGCGCTCGGC 3879  
QY 1600 GCGCGCTCCCGGAGGAGCGGCGGACACATCGGAGTACGGAGGACAGCTGTAC 1659  
Db 3880 GTCTCTGCTCTTCTCTCGGTGCGCGGATCTGGGCGGCGGCTGCGGCGCGCTAC 3939  
QY 1660 GTCGACGAGGAGCGCTCGGCTGCGGCTGCTGCGCGGCTGCTGCGGCGCGCG 1719  
Db 3940 GCGCGCTGAGCGCTTCTGACCGCGCTGCGGCTGAGCGCGGCGGCGGCGCGCG 3999  
QY 1720 GCGCGCTGAGCGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1779  
Db 4000 GTGACCTCGGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 4059  
QY 1780 GTCGAGGCTTTCGTAACCGCGGCGGCGGCTGCTTCCAGCAGGCTCCTTCCCGGAG 1839  
Db 4060 GAGCGGCTGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 4119  
QY 1840 GCGGACGAGGAACTCTGCTTACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1892  
Db 4120 GACCGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4172

## RESULT 10

US-09-861-289-32

; Sequence 32, Application US/09861289

; Patent No. US20020110897A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

```
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861.289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
us-09-861-289-32

Query Match      3.9%; Score 78.2; DB 10; Length 11220;
Best Local Similarity 41.9%; Pred. No. 34e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

Qy 469 GCCTCGCTCCGCTGCTACCATGACGCCCCCGAGCGCTGCTGCGACCCCA 528
Db 2740 GACTCCCCCGCGTCCAGGGCTCGGTGACGACTCCTGGCGCTACCGGATCGACTGGAAG 2799
Qy 529 GCCCGGTGACACACCGCGGCTTACCAGCTGTACTACCTGTGACTCCGACACAGAAC 588
Db 2800 GCCTCGGGTTCGCGAGCGCTCGAGCGCGCGGCTGTCGGGGCTGCTGCTGCTC 2859
Qy 589 GCGCCCGCGGCTGGGACACCGGAGCAGCAGCGCGCTGCTGCTTACGACACACGCGC 648
Db 2860 GTCCCGAGGACCGTTCGCGCGAGCGCGCGCGTGTGCTGCGCGCTGCTGCGCGCGC 2919
Qy 649 ACCGTGATCGCGTTCGCGCGCGCTTCCCGTGTGTCGGGTGCGGGTGTGTCGGGAC 708
Db 2920 GCCACCGCGTGTGAGCTGTGAGCTGTGCGCGCTGCGCGACCGCGAGCGGCTGCGCGG 2979
Qy 709 CGGAACACGGC---AGGTTTCGCGCGCGCGCTGCTGCGCTGCGGACCGCGGAC 765
Db 2980 CTGGCGAGGCGCTGGCGCGCGCGCTGAGCGCTGTGAGCGCTGTGAGCGCTGCTGCTGCG 3039
Qy 766 GACGGCTGCGCAAGTACAGGAGCAGTACTTCTACTGCTGCGAGCGCGGGTTCACG 825
Db 3040 TGGAGCAGAGCGCGCACCGCGCGCACCGCGCGCTTACCGCGGCGCACCGCGCGCAC 3099
Qy 826 TTCACCGCGCTGCGCGACCGCGCTATGCTCAACACCGAGGTGCGCGCGCGCACCGCGC 885
Db 3100 CTCACCTGTGCTGAGCGCTGAGAGCGCGCGCGCTGCGCGCGCGCTGTGCTGCTGAC 3159
Qy 886 CGCGAGATCGAGAGCGCGAGTGTGCTCGCGACCGCGCGCTGCTGCTGCTGCTGCTGCT 945
Db 3160 CACGGCGGCTGCTGCTGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3219
Qy 946 GGAGATGGGTCTCGCTATCGGACGACTGCGGTACGCGCGCTTCTACACCTCGCGCAAC 1005
Db 3220 TGGGCGATGGCGCGCTGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCTGCTGCT 3279
Qy 1006 CTGCGCGACTGACACTTCGCGCGCACTTCACTACCGAGCAGCGCTGCTGCGCGGATC 1065
Db 3280 CTGCGCTGCGAGCGCGCGCGCGCTGAGCGCGCTGAGCGCGCTGCTGCGCGCGGCT 3339
Qy 1066 GAGTGCCTCGGCTGCTGAGATCAACGCGACGAGCGGACCGCGCTGCTGCTGCTGCTGCT 1125
Db 3340 ACGGCTGAGGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3399
Qy 1126 GCCAGCATGGAGCGCTACGGCATCGGCTTCCCGATGAGCTACGCTACTGAGACGCGAC 1185
Db 3400 GCCTCCCTCCCGCGCGACGCGCGCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3459
Qy 1186 TGGAGCGGAGGATTCACGCGCGAGCAGCTACCGCGCGAATGGCTGCTGCTGCTGCTGCT 1245
Db 3460 GTCACCGGTGCTGAGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3519
Qy 1246 GACTGTGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
Db 3520 GCCGAGACCTTCTCTCTCCACACCGCGCTTCCGGCAGCGAAGCGCGCGCGCGCGCGCTCC 3579

; Sequence 5, Application US/09860846
; Patent No. US2002016472A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; CURRENT FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
us-09-860-846-5

Query Match      3.9%; Score 78.2; DB 9; Length 36778;
Best Local Similarity 41.9%; Pred. No. 3e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

Qy 469 GCCTCGCTCCGCTGCTTACCATGACGCCCCCGAGCGCTGCTGCGACCCCA 528
Db 18427 GACTCCCCCGCGTCCAGGGCTCGGTGAGGACTCTTGGCGCTACCGGATGCTGGAAG 18486
Qy 529 CGCGCGGTACACCGCGCGCGCTTACCGCTTACTACTGTACTGCTGCTGCTGCTGCTGCTGCT 588
```

Db 18487 GCGCTCGGGTCCGACGCGTCCGAGCGCGCGGGCTGTCGGGGCGTTCGTCGTC 18546  
Qy 589 GCGCCCGGGGCTGGGACACCGGAGCAGACGAGCGCGCTGCCCTTACGACACACCGC 648  
Db 18547 GTCCCGGAGACCGTTCGCGCGAGCGCGCGCGGCTGTCGCGCGCGTTCGCGCGCGC 18606  
Qy 649 ACCGTGATGCGCTGCGGGCCGACTTCCCGTGTGGTTCGGGTGCGGGTTCGCGCGAC 708  
Db 18607 GCGACCCCGGTACAGTGGACGTGTCGCGGTGGGAGCGCGGCTGCGCGCGAG 18666  
Qy 709 GGAACACGCG---AGGTTGCGCGCGCGCGGCTGTCGCGCTGCGGACACGCGGAC 765  
Db 18667 CTGGCGAGGGCCCTGGCGGGCGCGGTGAGCGCTGCGAGCGGGTCTTCTGCTGCTCG 18726  
Qy 766 GACGCGCTCCGCAAGTACAGGAGCAGTACTTACTGTGACGAGCGGGGTTCAGC 825  
Db 18727 TGGGAGGAGCGCGACCGCGCGCACCGCGCGCTTACCGGGGACCGCGCGAC 18786  
Qy 826 TTCACCGCCCTCGCGACCGCGCTCATGTCACACGAGTTCGCGCGCGCACCGCGC 885  
Db 18787 CTCACCTGTTGACGCGCTGAGGACCGCGCGTTCGCGCGCGCTGTGTGCTGAC 18846  
Qy 886 GCGGAGATCGAGACCGCGAGTGTTCGCGGACCGCAAGATCCACTGGGACACCGCGC 945  
Db 18847 CACGCGCGGTGTCGTCGCGCGCGCGCGACCGTCACTTCCCGCGCGCGCGCATGTG 18906  
Qy 946 GAGAAATGGGTGCTGCTCATCGGAGCTTGGGTGACGCGCGCTTACACCTTCGCGGAC 1005  
Db 18907 TGGGATGATGGCGGTGCGCGCTTGAGACACCGCGAGCGGTGGCGCGCTGATCAG 18966  
Qy 1006 CTGCGCGACTGAGACTTTCGCGCGCACTTTCGACTACCGGACCGCGCTTCGGCGGATC 1065  
Db 18967 CTGCGCTCGGACCGCGCGCGCGCGCTTGGACCGCATGACACCGTCTCGCGCGGT 19026  
Qy 1066 GAGTCCCGCGACTGTCGAGATCACCGACGAGCGGACGCGCACTGGTGTCTGCG 1125  
Db 19027 ACGGTGAGGACCGTTCGCGTACCGCTTCGCGGCTGCTCGCGCGCGCTTCGTCG 19086  
Qy 1126 GCCAGATGAGCGCTACGCGATCGCGCTCCCGATGAGTACGCTACTGGACGAGCGAC 1185  
Db 19087 GCTCCCTCCCGCGCGCGCGCGCTTTCGCGGTGGGAGCGCGCGCGCGCGCGTGTG 19146  
Qy 1186 TGGGACGCGGAGCTTCCACCGCGCGCGCTACCGCGCGCGCTGCTGCTGCGCGTGG 1245  
Db 19147 GTCACCGGTGCGGAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19206  
Qy 1246 GACTGTACGCGCGCTCACCTGGCGATCGATCGACGCGCGCGCGCGCGCGCGCG 1305  
Db 19207 GCGGACACTCTCTCTCACACCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 19266  
Qy 1306 ATCGGTGATGAACACTGGAAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1365  
Db 19267 GTGCGCGCGGAGCTTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19326  
Qy 1366 GGTACACGCGCGAGACTGATGCTGCGCGAGTTCGCGCGCGCGCGCGCGCGCG 1419  
Db 19327 ACGGCGCGGTGTCGATGCGCGCTACGACGCGCGCGCGCGCGCGCGCGCGCGCG 19386  
Qy 1420 GCGCGCTGTACACCGCTCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1479  
Db 19387 GCGCTCTCGAGCGCGCGCTGAGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 19446  
Qy 1480 ACCACACACTTCG 1539  
Db 19447 GAGCGCTTCG 19506  
Qy 1540 GCATACGAGATCGAGTGCATGCGCTGGGACCGCGCGCGCGCGCGCGCGCGCGCG 1599  
Db 19507 GCGCGCTTCACCTGAGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19566  
Qy 1600 GCGCGCTTCG 1659

Db 19567 GTCTGCTGCTCTTCTCTCTGTCGCGCGGATCTGGGGGGCGCGGTACAGGCGGTAC 19626  
Qy 1660 GTGACGAGGAGACCTCGACCTCGCGGGTACTCGCTGCGCGCGCTTACTGCGGAGCGCG 1719  
Db 19627 GCGCGCGGTACGCGCTTCTCTGACGCGCTTCTGCGCGGTTCAGCAGCGCGCGCGCG 19686  
Qy 1720 GCGCGCATCGACCGCGCGCGCGCTCGTGACCTTCTGCGACCTCTGTCGACACCGAGAC 1779  
Db 19687 GTGACCTCGGTGCGCTGAGCGCGCTGGGAGGCGACCGCGCTCACGAGGCTCGGACCG 19746  
Qy 1780 GTGAGGTCTTCTGTCACGCGCGCGCGCGCTGCTCTCCAGCAGGTCCACTTCCCGGAG 1839  
Db 19747 GAGCGCTGCGCGCGCTGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCTCACCGCGCTG 19806  
Qy 1840 GCGACACGGAATCTGCTCTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1892  
Db 19807 GACACGCGCTCGCGCGCGCGCGCGCGCTCACGATCGCGCGCGCTGCGACTG 19859  
RESULT 12  
US-09-861-289-5  
; Sequence 5, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-5

Query Match 3.9%; Score 78.2; DB 10; Length 36778;  
Best Local Similarity 41.9%; Pred. No. 3e-09;  
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;  
Qy 469 GCGTCTGCTGCGTTCACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528  
Db 18427 GACTCCCGCGCTCAGGGCTCCGTGCGAGACTCTTGGCGCTACCGCATCGACTGGA 18486  
Qy 529 GCGCGGTTCACACCGCGCGCTTACAGCTGTACTACTGCTGCTCGACCGAGAACAC 588  
Db 18487 GCGCTGCGCGTCCGCGCGCGTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTC 18546  
Qy 589 GCGCGCGCGCTGGACGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648  
Db 18547 GTCCCGGAGGACCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18606  
Qy 649 ACCGTGATGCGCGCTGCGCGCGCGCGCTTCCCGGTGCTGCGGTGCGCGGTCTGTCG 708  
Db 18607 GCGGACCGGTACAGTGGAGCTGTCGCGGTGGGCGAGCGCGCGCGCGCGCGCGCG 18666  
Qy 709 GCGAACACGCG---AGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 765  
Db 18667 CTGGCGAGGCGCTGCG 18726  
Qy 766 GACGCGTCCCAAGTACAGGAGCAGTACTTACTGCTGACCGCGCGCGGTTCAGC 825  
Db 18727 TGGGAGGAGCG 18786  
Qy 826 TTCACCGCGCTCG 885  
Db 18787 CTCACCTGCTGCG 18846

Qy 886 GCCGAGATCGAAGCGCGAGTGGTTCCGCGACCCACAGATCCACTGGGACACCGCCGC 945  
 Db 18847 CACGCGCGGTGTCTCGTCGGCGCGGCGGACACAGCTACCTCCCGCGCCAGGCGATGGTG 18906  
 Qy 946 GGAGATGGGTCTCGGTCTATCGGAGGACTGGGTACGGCGGGTTCATACACCTCGCGGAC 1005  
 Db 18907 TGGGCGATGGGCGGGTCCGCGCTCGGAGACCCCGAGCGGTGGGGCGGCTGATCGAC 18966  
 Qy 1006 CTGGCGAGTGGACACTTCGCGCGCAACTTCGACTACCCGGAACACAGCCCTCGGGGCGATC 1065  
 Db 18967 CTGGCTCGGAGCGGACCGGGCGGCTGGACCGCATGACACCGTCTCCCGGGCGGT 19026  
 Qy 1066 GAGTCCCCCGACCTGTTTCAGATCACCGAGACGAGCGGACACGCTGGGTGCTCGCC 1125  
 Db 19027 ACGGTGAGGACCAAGTTCGCGGTACGCGCTTCGCGGCTGCTCGCCCGCGCTCGTCCGC 19086  
 Qy 1126 GCCAGCATGGAGCGCTACGGCATCGGCTCCCGCATGAGCTAGCGTACTGACAGGCGAC 1185  
 Db 19087 GCGTCCCTCCCGGGCGACCGGACCGGCTTCGCGGTGGCGAGCGGACGCGACGCTGCTC 19146  
 Qy 1186 TGGAGCGCGAGCACTTCACCGCGGACGACCTCACCCCGCAATGGCTCGACTGGGGCTGG 1245  
 Db 19147 GTCACCGTGGCGAGGACCTTGGGCGCGCGAGGCGCGACGGCGTGGCGCGGACGCG 19206  
 Qy 1246 GACTGGTACGGCGGCGTCACTTGGCCATCGATCGACGCGCGCGGAGACCAAGCGCTCGCC 1305  
 Db 19207 GCGGACACCTCTCTCCACACACCCCTCCGCGAGGAGCGCGGAGGCGGCGGACCTCC 19266  
 Qy 1306 ATCGGTGGATGAACAATGGAAGTACCGCGACGCGAGCTCCCGACGAGCATCCGAC 1365  
 Db 19267 GGTGCGCGCGGAGACTCGGGGCTCGCGGGCTCGCTCGCGCAACTCGCGGAGCTGGGGCG 19326  
 Qy 1366 GCTTACAAAGCGGCAACTCGATCGTCGCGAGCTGGCGGCTCGCGCGGACGCTCT 1419  
 Db 19327 ACGGCGCGTGTGACTCGGACCTACGAGCGGAGGCGGCGCGCGGCTGCTCGCC 19386  
 Qy 1420 GCGGCTGGTACACCTCTTACGACCGCGGCTGGGAGCGGTGAGCACTACCTACCGCC 1479  
 Db 19387 GCGGCTCGGAGCGGACCGCTAGCGCGCTGCTCCACCTGCGCGCGGCGGCGGCTGCT 19446  
 Qy 1480 ACCACCACTCCCGGACCGGAGCTCGACGCGAGCGCGGCTGCTGCGATGAGACGAGCG 1539  
 Db 19447 GAGCGCTCGCGGAGCGAGCGGAGCGGCTCGCGCGGCTGCTGCGGAGCGGCGGCG 19506  
 Qy 1540 GCATACGAGATCGAGCTCGACATCGCTGGGACACCGGAGCGGAGCTCGGATCTCGGTG 1599  
 Db 19507 GCGGCGCTCGAGTGGACCGCTCTGCGGAGGCGCGGCTGCGGAGGCGGCTCGCGCC 19566  
 Qy 1600 GCGCGCTCCCGGAGCGGACCGGCGACACGACATCGGCAAGTACGAGCGGAGCGGCTGAC 1659  
 Db 19567 GTCTGTGCTCTTCTCTCGTGGCGCGATCTGGGCGGCGCGGCTGAGGCGGCTGAC 19626  
 Qy 1660 GTGACCGGAGGAGCTCGGAGCTCGCGGGTACTGCTGCGCGCGCTACTCGGAGCGCGC 1719  
 Db 19627 GCGCGGTAGCGGCTTCTCGAGCGGCTCGCGGCTGAGCGGCGGCGGCGGCGGCGGCG 19686  
 Qy 1720 GCGGCGATCGAGCGGCGGCGGATCGGTGACCTGCGCATCTCGTGCAGACCGAGCG 1779  
 Db 19687 GTGACCTGGTGGCTTGAGCGGCTTGGAGGCGGAGCGGCGGCTACCGAGGGTTCGACCGG 19746  
 Qy 1780 GTCGAGGTCTTCTGACCGGCGGCGACACCGTGTCTTCCGAGCGGCTGCTGCGCGAG 1839  
 Db 19747 GAGCGGCTGCGCGGCTTGGCGGCTGCGCGGCTGCGCGGCGGCGGCTGCGCGGCTG 19806  
 Qy 1840 GCGGACGCGGAGTCTGCTCTACACGAGCGGCGGCGGCGGCGGCGGCGGCTTACCGCG 1892  
 Db 19807 GACACCGGCTCGGCGGAGCGGAGACCGCGCTCAGATCGCGGAGCTGCGACTG 19859

RESULT 13

US-09-988-384B-5

; Sequence 5, Application US/09988384B

Publication No. US20030073824A1

GENERAL INFORMATION:  
 APPLICANT: Sherman, D. H.  
 APPLICANT: Liu, H.  
 APPLICANT: Xue, Y.  
 APPLICANT: Zhao, L.  
 TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 FILE REFERENCE: 600.536US1  
 CURRENT APPLICATION NUMBER: US/09/988.384B  
 PCT/US99/14398  
 PRIOR FILING DATE: 2001-11-19  
 PRIOR APPLICATION NUMBER: PCT/US99/14398  
 PRIOR FILING DATE: 1999-06-25  
 PRIOR APPLICATION NUMBER: US 09/105,537  
 PRIOR FILING DATE: 1998-06-26  
 NUMBER OF SEQ ID NOS: 53  
 SEQ ID NO 5  
 LENGTH: 37948  
 TYPE: DNA  
 ORGANISM: Streptomyces venezuelae  
 US-09-988-384B-5

Query Match 3.98; Score 78.2; DB 9; Length 37948;  
 Best Local Similarity 41.98; Pred. No. 3e-09;  
 Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;  
 Qy 469 GGCTCGCTCCGTGGCGTCTACCATGACGCGCCCGCCAGCGGTGGCTCTGGGACCGCCCAA 528  
 Db 19597 GACTCCCCCGCGTCCAGGGCTCGGTGAGGACTCTCTGGCGCTACCGCATCGACTGGAAG 19656  
 Qy 529 GCGCGGTACACACCGGCGGCTACGAGCTGTACTACGTGCTGCTGCGAGCAGACAC 588  
 Db 19657 GCGCTCGCGGTGCGCGACGCGTCCGAGCGCGCGGTGTCCGGCGGTGGCTGCTGCTG 19716  
 Qy 589 GCGCGCGCGGTGGGACACGCGAGCAGCAGCGGCGGTGCTGCGCGCGGTGCTGCGCGCG 648  
 Db 19717 GTCCCGGAGGACGTTCCGCGAGGCGCGCGGTGCTGCGCGCGGTGCTGCGCGCGCG 19776  
 Qy 649 ACCGTGATCGCGCTCGCGCGCGCTTCCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 708  
 Db 19777 GCGGACCGGTACAGCTGGAGCTGTCCCGGTGGGCGACCGGCGGCGCTCGCGCGGCG 19836  
 Qy 709 GCGACACGCG---AGGTTGCGCGCGCGGCGGCTGCTGCGCGCTGCGCGCGCGCGCG 765  
 Db 19837 CTGGCGAGGCGGCTTGGCGCGCGCGGCGGCGGTGAGCGGCTGCGCGCGGTGCTGCTG 19896  
 Qy 766 GAGCGCGTCCGCAAGTACAGGAGCAGTACCTTACTGCTGACGCGGCGGCTTACCG 825  
 Db 19897 TGGGACGAGAGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19956  
 Qy 826 TTCACCGCGCTCGCGCGCGCGCTCATCTCAACACCGAGGTCGCGCGCGCGCGCGCGCG 885  
 Db 19957 CTCACCGTGGTGCAGGCGCTGGAGGACCGCGCGCTGCGCGCGCGCGCGCGCGCG 20016  
 Qy 886 GCGGAGATCGAGACGCGGAGTGGTTCGCGACCGCGCGCGCGCGCGCGCGCGCGCG 945  
 Db 20017 CACGCGCGGTGCTCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20076  
 Qy 946 GGAGATGGGTCTCGTTCATCGGAGCTGCGGTACGCGCGGTGCTTACACCTCGCGCGAC 1005  
 Db 20077 TGGGCGATGGGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20136  
 Qy 1006 CTGGCGGACTGGACACTTCGCGCGCAACTTCGACTACCGCGCGCGCGCGCGCGCGCG 1065  
 Db 20137 CTGGCGCTCGGAGCG 20196  
 Qy 1066 GAGTCCCCCGGAGCTGTTTCAGATCACCGCAGACGAGCGGACACGCGCTGCTGCGCG 1125  
 Db 20197 ACGGTGAGGACAGGTGCGGCTACGCGCTCGCGGTGCTCGCGCGCGCGCGCGCGCG 20256  
 Qy 1126 GCCAGCATGGAGCGCTACGGCATCGGCGCTCCCGCATGAGCTAGCGCTACTGACAGGCG 1185  
 Db 20257 GCTTCCCTCCCGGCGGACGCGGACGCGCTTTCGCGGTGGTGGCGGCGGCGGCGCGCT 20316

QY 1186 TGGACGGGAGAGTTCCACCGCAGACCTCACCCCGCAATGCTCGACTGGGCTGG 1245  
Db 20317 GTACCGGTGCGGAGAGCTTCGGCGCGGAGCGGACGCGGCTGGCCCGCAGCGC 20376  
QY 1246 GACTGGTACGGCGGTACCTTGGCCATGATGATGAGGCGCCGAGACCAAGCGCTGGC 1305  
Db 20377 GCGGACACCTCTCTCTCACACCAACCCCTCCGGCAGCAAGGCGCGAAGGCACCTCC 20436  
QY 1306 ATTCGGTGGATGAACAACCTTGAAGTACGCGCAGCGAGTCCCGACCGACGACATCCGAC 1365  
Db 20437 GGTGCGCGGAGACTTCGGCCCTCGCGGCTGCTGCGGAACTCGCGGACCTGGGCGG 20496  
QY 1366 GCTTAAACGGGAGAACTCGATGTCGCGAGCTGCGGCTGCGCCGACAGC-----CT 1419  
Db 20497 ACGGCAACGCTGTGACCTGCGACCTACGGACGCGGAGGCGCGCGGCTGCTCGGC 20556  
QY 1420 GCGGCTGGTACACCTCTGTAGGACCCCGCTGGGAGCGCTGACAACTACGTCACCGGC 1479  
Db 20557 GCGGCTCGGAGCGACCCGCTCAGCGCGCTCTCCACCTGCGCGCCACCGTGCACCTCC 20616  
QY 1480 ACCACACACTCCCGGACCGGACCGCTGACGCGAGCGGCTCTGCCATGGAACGAGGC 1539  
Db 20617 GAGCGCTCGCGCGACCGGACGCGCTCGCGCTGCTGACCGCGAAGCCACC 20676  
QY 1540 GCATACGAGATCGAGCTGCATGTCGCGAGCTGGGACACCGGAGAACTGCGGCTCGGTG 1599  
Db 20677 GCGGCTGCTGAGCGGCTGCTGCGGAGCGGCTGCGGAGCGGCTGCGGAGCGGCTCGGCC 20736  
QY 1600 GCGCGCTCCCGGAGGACCGGACGACGACGACATCGCAATCGGACGACGACCTGTAC 1659  
Db 20737 GTCTGTGCTCTTCTCTGCTGCGCGGATCTGGGCGGCGGCTGAGGCGCGGTAC 20796  
QY 1660 GTGACCGGAGGACCTCGGACCTCGCGGCTGCTGCGCGGCTGCTGCGGAGCGCGC 1719  
Db 20797 GCGCGGTGAGCGCTTCTCGACGCTCTCGCGCTGACGACCGGCGGCGGCGCGCACC 20856  
QY 1720 GCGCGCTGAGCGCGGCGCGGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1779  
Db 20857 GTGACCTGCGGTGAGCGGCTGCGGAGCGGCTGCGGAGCGGCTGCGGAGCGGCTGCGG 20916  
QY 1780 GTGAGGTCTTGTCAACCGCGGCGGACGCGGCTGCTGCGGAGCGGCTGCGGAGCGG 1839  
Db 20917 GAGCGGCTGCGGCGCTGCGGCTGCGGCGGCTGCGGCGGCGGCGGCTGCGGCGGCTG 20976  
QY 1840 GCGGACCGGGAATCTGCTTACACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1892  
Db 20977 GACACCGGCTCGGCGCACCGGACACCGCGCTCAGGATCGCGGAGCTCGACTG 21029

## RESULT 14

US-09-988-384B-3/c  
; Sequence 3, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600,536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 3  
; LENGTH: 12441  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae

US-09-988-384B-3

Query Match 3.9%; Score 77.6; DB 9; Length 12441;  
Best Local Similarity 42.9%; Pred. No. 4.6e-09;  
Matches 680; Conservative 0; Mismatches 879; Indels 25; Gaps 5;  
QY 367 CCGGCCATCTCAGCGCGGCGGCTGCTCAGGAGGAGCGGCGGAGGACTCGCCCTGATC 426  
Db 11836 CCGGACACCGCAACTCTGACCGGACGAGGTGGCGCGCGGTCCACACCGCGACCTCG 11777  
QY 427 TTCGCGGTGTGTCGCGCTTGCACGCGGCGATCCGCTCGCGGCTCGCTCGCTCGCTC 486  
Db 11776 GCGGCTGCGGCTGCTGCTGCGGCGGCGCTCGCGGCGGCGCTCGCTCGCTCGCTC 11717  
QY 487 TACCACATGAGCGCGCGCGGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 546  
Db 11716 GCGGACGAGCAGCGCTGCGGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 11657  
QY 547 GCGGCTTACGAGCTTACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606  
Db 11656 GAGGCGCGCGCGCGCGGAGCGCTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 11597  
QY 607 CAGCGAGCAGCAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666  
Db 11596 GCGGCTTACGAGCTTACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11537  
QY 667 CCGGCTTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726  
Db 11536 ATCCGCGCGCTTCCACAACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11477  
QY 727 GCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 786  
Db 11476 AAGCGAAGATGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11429  
QY 787 GAGCAGTACCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 846  
Db 11428 GAGCGGCTTCCGAGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11369  
QY 847 GTATGCTGAACACGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 906  
Db 11368 CTCGCGGAGCTTCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11309  
QY 907 TGGTTCGCGGACCGCAAGATCCACTGCGGACACCGCGCGGAGATGGCTGCTGCTGCT 966  
Db 11308 CAGTACGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11249  
QY 967 GAGCAGCTCGGTACGCGCGCTTCTACACCTGCTGCGGAACT-----GCGGAGCTGGACA 1020  
Db 11248 GAGGCTCTGAGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11189  
QY 1021 CTTGCGCGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
Db 11188 CTGGAGCGGTATCCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11129  
QY 1081 TTGAGATCAGCGCAGAGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
Db 11128 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11069  
QY 1141 TACGCGATCGGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
Db 11068 GACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11009  
QY 1201 TTCCAGCGCGGAGCTTACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
Db 11008 GCG 10950  
QY 1261 GTACCTGGCGATGATGAGCGCGCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
Db 10949 ACCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10890  
QY 1321 AACTGGAAGTACGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
Db 10889 CTGGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10830  
QY 1381 AACTCGATCGTCCG---CGAGCTGCGGCTGCGCGCGGAGCGGCTGCTGCTGCTGCTGCT 1437









GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 02:37:06 ; Search time 5273.5 Seconds  
(without alignments)  
11042.896 Million cell updates/sec

Title: US-09-868-328B-3  
Perfect score: 2001  
Sequence: 1 gcggtgcaccccgactccc.....atcgacacgtctctcgctt 2001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match % | Score | Length | DB       | ID           | Description        |
|------------|---------------|-------|--------|----------|--------------|--------------------|
| 1          | 78.3          | 1566  | 1      | AF181254 | AF181254     | Arthrobac          |
| 2          | 78.3          | 1566  | 1      | AF294730 | AF294730     | Arthrobac          |
| 3          | 945.2         | 47.2  | 1752   | 1        | AB001984     | Arthrobac          |
| 4          | 931.8         | 46.6  | 1551   | 6        | E22826       | Levan fruct        |
| 5          | 909.6         | 45.5  | 1452   | 6        | E22825       | Levan fruct        |
| 6          | 370.4         | 18.5  | 1863   | 1        | AB062062     | Microbacte         |
| 7          | 147.6         | 7.4   | 65140  | 6        | AX211705     | Sequence           |
| 8          | 147.6         | 7.4   | 123580 | 1        | AF263912     | Streptomy          |
| 9          | 147.6         | 7.4   | 125401 | 6        | AX211739     | Streptomy          |
| 10         | 141.8         | 7.1   | 3030   | 8        | VCA429230    | Volvox ca          |
| 11         | 119.8         | 6.0   | 3849   | 6        | AX058889     | Sequence           |
| 12         | 119.8         | 6.0   | 12744  | 1        | AF204401     | Streptomy          |
| 13         | 114.2         | 5.7   | 22449  | 1        | SPSNBCDE     | Y11548 S.pristinae |
| 14         | 114.2         | 5.7   | 22449  | 1        | SPSNBCGEN    | X98690 S.pristinae |
| 15         | 112           | 5.6   | 77457  | 1        | AF210249     | Streptomy          |
| 16         | 111.8         | 5.6   | 2951   | 1        | AF123319     | Streptomy          |
| 17         | 111.8         | 5.6   | 43147  | 1        | SC4A10       | Streptomy          |
| 18         | 109.2         | 5.5   | 27541  | 6        | AX211706     | Sequence           |
| 19         | 106.8         | 5.3   | 30000  | 6        | AX250262     | Sequence           |
| 20         | 106.8         | 5.3   | 88063  | 1        | BSZ94043     | B.subtilis         |
| 21         | 106.8         | 5.3   | 209510 | 1        | BSUB0018     | 299121 Bacillus su |
| 22         | 106.4         | 5.3   | 113193 | 1        | AF357202     | Streptomy          |
| 23         | 105.2         | 5.3   | 13424  | 1        | AE007019     | Mycobacte          |
| 24         | 105.2         | 5.3   | 58930  | 2        | AC098321     | Rattus no          |
| 25         | 103.8         | 5.2   | 47852  | 1        | MTV023       | Streptomy          |
| 26         | 103.6         | 5.2   | 14860  | 1        | AE007093     | Mycobacte          |
| 27         | 103.6         | 5.2   | 63033  | 1        | MTV008       | Mycobacte          |
| 28         | 102.4         | 5.1   | 303091 | 2        | AC084799     | Mus muscu          |
| 29         | 102.2         | 5.1   | 298166 | 2        | AC087563     | Homo sapi          |
| 30         | 101.6         | 5.1   | 5994   | 1        | SVSNBDE      | Y11547 S.virginiae |
| 31         | 101.4         | 5.1   | 42655  | 1        | SC7H2        | Streptomy          |
| 32         | 101           | 5.0   | 15311  | 1        | AE007164     | Mycobacte          |
| 33         | 99.8          | 5.0   | 4068   | 1        | ANU12274     | Actinomyc          |
| 34         | 99.8          | 5.0   | 37245  | 1        | SC5F2A       | Streptomy          |
| 35         | 99.8          | 5.0   | 40790  | 1        | MTCY493      | Mycobacteri        |
| 36         | 99.4          | 5.0   | 220469 | 2        | AC074307     | Mus muscu          |
| 37         | 99            | 4.9   | 15348  | 1        | AE007163     | Mycobacte          |
| 38         | 98.6          | 4.9   | 67200  | 1        | MTV017       | Mycobacte          |
| 39         | 98.6          | 4.9   | 77534  | 1        | AF235504     | Streptomy          |
| 40         | 98.2          | 4.9   | 110000 | 2        | LMFLCHR32_06 | Continuation (7 of |
| 41         | 97.8          | 4.9   | 34108  | 1        | SCE25        | Streptomy          |
| 42         | 97.6          | 4.9   | 150754 | 9        | AC023491     | Streptomy          |
| 43         | 97.2          | 4.9   | 10420  | 1        | AE005127     | Homo sapi          |
| 44         | 97            | 4.8   | 4615   | 1        | STMWH12X     | Halobacte          |
| 45         | 96.6          | 4.8   | 46166  | 1        | SC8F11       | Streptomy          |
|            |               |       |        |          |              | AL353864           |

ALIGNMENTS

RESULT 1  
AF181254  
LOCUS AF181254  
DEFINITION Arthrobacter ureafaciens levan fructotransferase (lfta) gene,  
complete cds.  
ACCESSION AF181254  
VERSION AF181254.1 GI:8163731  
KEYWORDS Arthrobacter ureafaciens.  
SOURCE Arthrobacter ureafaciens  
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.  
REFERENCE 1 (bases 1 to 1566)  
AUTHORS Song,K.B., Rhee,S.K. and Yoo,E.J.  
TITLE Nucleotide sequence of levan fructotransferase gene (lfta) from

```
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
SUBMITTED (27-AUG-1999) Microbial Metabolic Engineering RU, Korea
Research Institute of Bioscience and Biotechnology, Taejeon 305-600,
South Korea
FEATURES
source
1. .1566
Location/Qualifiers
/organism="Arthrobacter ureafaciens"
/strain="K-2032"
/db_xref="taxon:37931"
/country="South Korea"
1. .1566
/feature="lftA"
1. .1566
/feature="lftA"
/function="diffructose dianhydride IV (DFA IV) formation
from levan"
/feature="levan-assimilating enzyme"
/codon_start=1
/transl_table=11
/product="levan fructotransferase"
/protein_id="AAF73829.1"
/db_xref="GI:8163732"
/translation="MPLSRRAVVLQAGAGALALIFGAVPPAARASAPGSLRAYVH
MTPPSGLCDPQRYTHGAYQLYLLHSDNNGPGMDHASTIDGVAFTHRGTVMPLR
PDPFVMSAVVGTANTAGFAGAVVALATOPTDGVRYQYQYLYWSDGFTFALP
DPVINTDGRYPAETIENAEWFRDPIKHWDTARGWVCVIGRLRYAAFTVSPNLRD
WTLRNFDPNHALGGIECPDLFEITADDGPRHWLAASMDAYGLIGLPMYVWTGTW
DGEQPHADLLPQWLWDGWDYAAVTFESIDAPEFKRLAIAMNNWKYAKRDVPTDAS
DGYNSQSLVRELRLARQPGWYLLSPVAALTNVATTTLPDRTVDGSAVLDPWG
RAYETIEDIANTATNVGIVSVPDGTNRHIGYADLYVDRGSPSLAGYSLAPYS
RAAAPIDPGARSVHLRLVDTQSVFVFNAGHTVLSQOVHFAEGDTGISLTDGPAH
FTGIIVVREIGQAI"
BASE COUNT 268 a 635 c 458 g 205 t
ORIGIN
Query Match 78.3%; Score 1566; DB 1; Length 1566;
Best Local Similarity 100.0%; Pred. No. 1.1e-160;
Matches 1566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 361 ATGACCGCGGCATCTACGCGCGCGCGTCTCCAGGAGCGCGCGCGGAGACTCGCC 420
DB 1 ATGACCGCGGCATCTACGCGCGCGCGTCTCCAGGAGCGCGCGCGGAGACTCGCC 60
QY 421 CTGATCTCGCGGCTCTGCGCGCTGCGCGCGCGGATCGCTCCGCGCTCGCT 480
DB 61 CTGATCTCGCGGCTCTGCGCGCTGCGCGCGCGGATCGCTCCGCGCTCGCT 120
QY 481 GCGGTCTACACATGACGCGCGCGCGCTGCGCGCGCGGATCGCTCCGCGCTCGCT 540
DB 121 GCGGTCTACACATGACGCGCGCGCGCTGCGCGCGCGGATCGCTCCGCGCTCGCT 180
QY 541 ACCGCGCGCGCTACGAGTGTACTACCTGCGCGCGCGGATCGCTCCGCGCGCGG 600
DB 181 ACCGCGCGCGCTACGAGTGTACTACCTGCGCGCGCGGATCGCTCCGCGCGCGG 240
QY 601 TGGGACACGCGGACGACGCGCGCTGCGCTTCCAGCGCGCGGATCGCTCGCGG 660
DB 241 TGGGACACGCGGACGACGCGCGCTGCGCTTCCAGCGCGCGGATCGCTCGCGG 300
QY 661 CTGCGCGCGGCTTCCCGGTGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 720
DB 301 CTGCGCGCGGCTTCCCGGTGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 360
QY 721 GGGTTCGCGCGCGCGGTGCGTTCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 780
DB 361 GGGTTCGCGCGCGCGGTGCGTTCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 420
QY 781 TACAGGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
```

```
DB 421 TACAGGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
QY 841 GACCCCGTTCATCTCAACACCGAGCGTTCGCGCGCGGATCGCGCGCGGATCGG 900
DB 481 GACCCCGTTCATCTCAACACCGAGCGTTCGCGCGCGGATCGCGCGCGGATCGG 540
QY 901 GCGGAGTGTTCGCGGACCGGATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
DB 541 GCGGAGTGTTCGCGGACCGGATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
QY 961 GTCATCGGAGCACTCGGTTACCGCGGTTCTACCTCTCGCGGAACTCGCGGATCG 1020
DB 601 GTCATCGGAGCACTCGGTTACCGCGGTTCTACCTCTCGCGGAACTCGCGGATCG 560
QY 1021 CTTGCGCGCAACTTCGACTTACCGGACCGGATCGCTCTCTCTCTCTCTCTCTCT 1080
DB 661 CTTGCGCGCAACTTCGACTTACCGGACCGGATCGCTCTCTCTCTCTCTCTCTCT 720
QY 1081 TTCGAGATCACGCGAGACCGGATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
DB 721 TTCGAGATCACGCGAGACCGGATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
QY 1141 TAGGCGATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
DB 781 TAGGCGATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
QY 1201 TTTACGCGCGAGCACTCTACCGCGGATCGCTCTCTCTCTCTCTCTCTCTCTCT 1260
DB 841 TTTACGCGCGAGCACTCTACCGCGGATCGCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 1261 GTCACTTGGCGATCGATCGCGCGCGGATCGCTCTCTCTCTCTCTCTCTCTCTCT 1320
DB 901 GTCACTTGGCGATCGATCGCGCGCGGATCGCTCTCTCTCTCTCTCTCTCTCTCT 960
QY 1321 AACTGGAAGTACGCGCGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
DB 961 AACTGGAAGTACGCGCGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
QY 1381 AACTCGATCTCGCGAGCTCGCGGCTCGCGGATCGCTCTCTCTCTCTCTCTCTCT 1440
DB 1021 AACTCGATCTCGCGAGCTCGCGGCTCGCGGATCGCTCTCTCTCTCTCTCTCTCT 1080
QY 1441 AGCAGCGCGTGGCGAGCTGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
DB 1081 AGCAGCGCGTGGCGAGCTGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
QY 1501 ACCGTGCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
DB 1141 ACCGTGCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
QY 1561 ATCGCTTGGGACACCGCGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
DB 1201 ATCGCTTGGGACACCGCGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
QY 1621 CGGCGACAGCAATCGGCAAGTACGAGGAGCACTCTCTCTCTCTCTCTCTCTCTCTCT 1680
DB 1261 CGGCGACAGCAATCGGCAAGTACGAGGAGCACTCTCTCTCTCTCTCTCTCTCTCTCT 1320
QY 1681 CTCTCGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
DB 1321 CTCTCGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
QY 1741 CGATCTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1800
DB 1381 CGATCTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
QY 1801 GCGCGACCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1860
DB 1441 GCGCGACCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
QY 1861 TACAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
DB 1501 TACAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
```





Db 1170 TCCGGCGCTTCGAACTATGCACCTCCAGCACCACTTCGGGACCCACAGTCAACGG 1229  
Qy 1512 CAGCCCGCTTCCTGCATCAAGGACGCGCATACAGAGATCAGGTTCGACATCGCTGGGA 1571  
Db 1230 CAGTTTCGACTTCGGTGGAGCGCGCGGTATGAAGTGAATCTGATATTTTCATGGGA 1289  
Qy 1572 CACCGGACGAAAGTCGCGATCTCGGTGGCGCGTCCCGCCGACGGAACCCGGCACACGAA 1631  
Db 1290 CAGGCGACGAAAGTGGGAGTCTCGGTGGCGCGTCCGATGCGATGCGACGCGCCATACGAA 1349  
Qy 1632 CATCGGCAAGTACGAGCAGACCTGTAGTTCAGCGGAGGACCTTCGACCTTCGCGGGTA 1691  
Db 1350 CATCGGCAATACGCTGACGAGTGTAGTTCGATCGCGCATCTTCGGAGCAAGCGGTA 1409  
Qy 1692 CTCGCTCCCGCTTACTCGCGAGCGCGCGCCCATGACGCCCGCGCGATCGGTGCA 1751  
Db 1410 TGGCGTGACCCCTACACCGCGCGCGCGCCCATGATCGAGCGGACATCGGTCCA 1469  
Qy 1752 CCGCGCATCTCGTCGACACCCAGAGCGTTCGAGTCTTCGTCGAACCGCGCGCACACCGT 1811  
Db 1470 CCGCGCATCTCTTAGACACCCAAAGTGTGAGGTGTTCGTAATTCGCGGCACACGGT 1529  
Qy 1812 GCTCTCCAGAGTCCACTTCGCGGAGCGCGGCGACACGGAATCTCGTCTACACGCGG 1871  
Db 1530 GGTTCGAGCAGGTGCACTTCGCGGCGCGGCGGACACGCGGATCTCCCTCTATGCGGAGG 1589  
Qy 1872 CGGCGCGCACATCAACCGCATCTCGTCCGCGAGATTCGCCAGGCGATCTAGGCGAT 1931  
Db 1590 CGGTCCGCGCAACTCAACCGCATCAACATCCGCGAGTTCGGGAACCCCATCTAAGCCTG 1649  
Qy 1932 GCACACACACCGTTCACCGAAGCGCGCGCGCGGAGACGACGG 1975  
Db 1650 CGTCCCGCGGTGGAAGGACGCGCGGACGCTGCAGCAGGCGG 1693

## RESULT 4

E22826  
LOCUS E22826 1551 bp DNA linear PAT 18-JUN-2001  
DEFINITION Levan fructotransferase gene.  
ACCESSION E22826  
VERSION E22826.1 GI:13024125  
KEYWORDS JP 1999069978-A/2.  
SOURCE Arthrobacter nicotinovorans.  
ORGANISM Arthrobacter nicotinovorans  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micrococciaceae; Micrococaceae; Arthrobacter.  
REFERENCE  
AUTHORS Masakazu, S., Atsushi, Y. and Fusao, T.  
TITLE Levan fructotransferase gene  
JOURNAL Patent: JP 1999069978-A 2 16-MAR-1999;  
NIPPON OIL CO LTD  
COMMENT OS Arthrobacter nicotinovorans  
PN JP 1999069978-A/2  
PD 16-MAR-1999  
PF 28-AUG-1997 JP 1997232421  
PR

PI MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TOMITA  
PC C12N15/09, C07K14/195, C12N1/21, C12N9/10//A61K38/45, (C12N15/09,  
PC C12R1:06),  
PC (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,  
PC (C12N15/00, C12R1:06)  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..1551  
FT Location/Qualifiers  
1..1551 /organism='Arthrobacter nicotinovorans',  
/organism='Arthrobacter nicotinovorans'  
/db\_xref='taxon:29320'

## FEATURES

source

## BASE COUNT

ORIGIN

315 a 524 c 443 g 269 t

Query Match 46.6%; Score 931.8; DB 6; Length 1551;  
Best Local Similarity 75.8%; Pred. No. 3.6e-92;  
Matches 1184; Conservative 0; Mismatches 367; Indels 12; Gaps 2;

Qy 361 ATGAGCGCGGCATCTCAGCGCGCGCGTGTCTCAGGAGAGCGCGCGGAGACTCGCC 420  
Db 1 ATGAGCTATGACATCTCTCGCGCGCATCTCCCTGCAAGGTGAGGGGTGGTGGCTTGGCA 60  
Qy 421 CTGATCTTCGGCGGTGTGTGCGCTGCGCGCGGCGATCCGCTCGGGGTCTCGCTCGGT 480  
Db 61 CTTTTCATGAGCAATGCCATTTCCTCGTGGCGCCCA-----CGCCAGGCAATCCCTCCGG 114  
Qy 481 GCCCTTACCATATGAGCGCGCGCGCGGTGTCTGCGACCCCGCCCAACGCGCGCTCAC 540  
Db 115 GCGATCTAGCACATGACCGCGCGCTCGGCTGGCTGTATGTATGATCGCGAGGACCGTACAT 174  
Qy 541 ACCACGCGCGCTACGAGCTGTACTACGTGACATCGACACAGACAGAGCGCGCGCGGCG 600  
Db 175 ACAACGCGCGCTACGAGCTCTACTCTCAGTCCGCGCCAGAAACAGCGACCGGGCGGA 234  
Qy 601 TGGACACGCGGACGACGCGCGGTGCGCTTTCAGCGACACGCGACCGCGTGTATGCGG 660  
Db 235 TGGACACGCGGACGACGCGCGGTGCGCTTTCAGCGACACGCGACCGCGTGTATGCGCA 294  
Qy 661 CTGCGCGCGCATCTCCCGGTGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 720  
Db 295 ATGCAACGCGACTTCCCGGTGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 354  
Qy 721 GGTTCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 780  
Db 355 GGTTCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 408  
Qy 781 TACAGGAGCAGTACTCTACTGTCGACGCGCGGTTCAGGTTCACGCGCTCGCC 840  
Db 409 TTCAGGAGCAGTACTCTACTGTCGACGCGGTTCAGGTTCACGCGCTCGCC 468  
Qy 841 GACCCCGTCACTCGTCAACACGCGGTGCGCGCGCGGTTCAGGTTCACGCGCTCGCC 900  
Db 469 GACCCCGTCACTCGTCAACACGCGGTGCGCGCGCGGTTCAGGTTCACGCGCTCGCC 528  
Qy 901 GCGAGTGGTTCGCGCGCGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 960  
Db 529 GCAGAAATGGTTCGCGCGCGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAG 588  
Qy 961 GTCATCGGACGACTGCGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGT 1020  
Db 589 GTCATCGGACGACTGCGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGTTCAGGT 648  
Qy 1021 CTTGCGCGCAACTTCGACTACCCGAAACACGCGCTTCGCGCGCATCGAGTTCGCGCGCAT 1080  
Db 649 TGAAGTCCAACTTCGACTACCCGAAACACGCGCTTCGCGCGCATCGAGTTCGCGCGCAT 708  
Qy 1081 TTCGAGATCACCGCAGGAGGACGCGCATCGGTTCGCGCGCATCGAGTTCGCGCGCATCGAG 1140  
Db 709 TTCGAGATCACCGCAGGAGGAGGACGCGCATCGGTTCGCGCGCATCGAGTTCGCGCGCAT 768  
Qy 1141 TACGCGCATCGCGCTTCGACTGACCTTACTGAGCAGGACGCGCTTCGCGCGCATCGAG 1200  
Db 769 TACGCGCATCGCGCTTCGACTGACCTTACTGAGCAGGACGCGCTTCGCGCGCATCGAG 828  
Qy 1201 TTCACGCGCGGACGACTTCACCGCGCATCGGTTCGCGCGCATCGAGTTCGCGCGCATCGAG 1260  
Db 829 TTCATCGCGGACGACTTCACCGCGCATCGGTTCGCGCGCATCGAGTTCGCGCGCATCGAG 888  
Qy 1261 GTCACCTTGGCGCATCGATCGAGCGCGCGCGGACGCGCATCGGTTCGCGCGCATCGAG 1320  
Db 889 GTGACCTTGGCGCGCGGTGGAAGCAGCTTCGCGCGCATCGGTTCGCGCGCATCGAG 948  
Qy 1321 AACTGGAAGTACCGCGCGCGGACGCGCATCGGTTCGCGCGCATCGGTTCGCGCGCATCGAG 1380  
Db 949 AACTGGAAGTATGCGCGCGCGGACGCGCATCGGTTCGCGCGCATCGGTTCGCGCGCATCGAG 1008

QY 1381 AACTCGATCGTCGGGAGCTGGGCTCGCCGACACGCTGGGGCTGTGTACACCTCTCTG 1440  
Db 1009 AATTCATCACCGAGCTAGGCTCGAGCGCAATCGGGGGTGTGTACACCTTGTCTC 1068  
QY 1441 AGCACCCCGTGGCAGCGTACGAACTACGTCACCGCCACACACACACTTCCCGACGG 1500  
Db 1069 AGCAGCCCGTGGGCGGTTCGAACTATGCCACTTCAGACACACACCTTCCGGACCGC 1128  
QY 1501 ACCGTGACGAGCGCCCTCTGTCATGGAACGAGCGCATAGAGATCGAGTTCGAC 1560  
Db 1129 ACAGTCAACGGCAGTTCGTACTTCGTGGAGCGCGCGGTATGAACCTGAACCTCGAT 1188  
QY 1561 ATCGCTGGGACACCGCAGAACGTGCGCATCTCGTGGGCGGTCTCCCGCAGCGAAC 1620  
Db 1189 ATTTATGGGACACGCGAACGTGGGAGTCTCGTGGGCGGTCTCGTCCGATGGCAGC 1248  
QY 1621 CGGCACACCAATCGGCAAGTACGAGCAGACCTGTAGTGCAGCGGAGGACCTCCGAC 1680  
Db 1249 CGCATACCAATCGGCAATACGTTGACGAGTGTAGTGCATCGCGCATCTCGGAG 1308  
QY 1681 CTCGCGGTACTCGTCGCCCTCTACTCGGAGCGCGCGCCCATCGACCCCGCGCC 1740  
Db 1309 CAAAGCGGTATGCTGCGCACCTTACACCGCGCGCGCGCCCATCGATCGAACGCC 1368  
QY 1741 CGATCGGTGACCTCGGATCTCTGTCGACACCGAGCGTGTCTTCTCAAGCGC 1800  
Db 1369 AGATCGGTGACCTCGGATCTCTTGTAGACACCAAGTGTGAGGTGTCTGTAATTC 1428  
QY 1801 GCGCACACCGTCTCTCCAGCAGTCCACTTCGCGAGCGCGGACACGGAATCTCGTC 1860  
Db 1429 GGGCACAGGTGTTGCGCAGAGTGCATTCGCGCGCGCGGACACGGGGATCTCCCTC 1488  
QY 1861 TACACGAGCGCGCGCCGACACTTACCGGATCTGTGTCGCGGAGATGGCCAGCGC 1920  
Db 1489 TATCGGAGCGCGGTTCGGCCCACTTACCGGCATCACCATCGCGAGTTCGGGAACCC 1548  
QY 1921 ATC 1923  
Db 1549 ATC 1551

RESULT 5  
E22825  
LOCUS E22825 1452 bp DNA linear PAT 18-JUN-2001  
DEFINITION Levan fructotransferase gene.  
ACCESSION E22825  
VERSION E22825.1 GI:13024124  
KEYWORDS JP 199069978-A/1.  
SOURCE Arthrobacter nicotinovorans.  
ORGANISM Arthrobacter nicotinovorans  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micrococciaceae; Micrococccaceae; Arthrobacter.  
REFERENCE  
AUTHORS Masakazu, S., Atsushi, Y. and Fusao, T.  
TITLE Levan fructotransferase gene  
JOURNAL Patent: JP 199069978-A 16-MAR-1999;  
NIPPON OIL CO LTD  
COMMENT OS Arthrobacter nicotinovorans  
PN JP 199069978-A/1  
PD 16-MAR-1999  
PF 28-AUG-1997 JP 1997232421  
PR  
PI MASAKAZU SAITO, ATSUSHI YOKOTA, FUSAO TOMITA  
PC C12N15/09, C07K14/195, C12N1/21, C12N9/10//A61K38/45, (C12N15/09,  
C12N1/06),  
PC (C12N1/21, C12R1:19), (C12N9/10, C12R1:19), C12N15/00, A61K37/52,  
PC (C12N15/00, C12R1:06)  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1. .1452  
FT /organism='Arthrobacter nicotinovorans'.  
FT Location/Qualifiers

source 1. .1452  
/organism='Arthrobacter nicotinovorans'  
/db\_xref='taxon:29320'  
BASE COUNT 299 a 491 c 416 g 246 t  
ORIGIN  
Query Match 45.5%; Score 909.6; DB 6; Length 1452;  
Best Local Similarity 77.0%; Pred. No. 9.3e-90;  
Matches 1123; Conservative 0; Mismatches 329; Indels 6; Gaps 1;  
QY 466 CGGGCTCGCTCGGTCGGTCTACACATGACGCCCGCAGCGGTCTGTGCGACCCC 525  
Db 1 CAGGATCTCCCTCGGGCGATCTACACATGACCCCGCCCTCGGGCTGGTATGTATCG 60  
QY 526 CAACCGCGGTCAACACCCACCGCGCTTACAGCTGTACTAGCTGCATCTCCGACGAA 585  
Db 61 CAGCGCCGTATACATAACAACGCGCTTACAGCTTACTTCCACTCCGCGCCAGAAC 120  
QY 586 AACGCGCGCGGTGGGACACCGGAGCAGCGGCGGTCTGCTTACGACGACAC 645  
Db 121 AACGGACCGGGATGGGACACCGGACGACCGGCGAGTGTCTTTACACCCACCAT 180  
QY 646 GGCACCGTATCGCGCTGCGGCCCGACTTCCCGGTGTGTGCGGTGCGCGTCTGCGC 705  
Db 181 GGATGTGTATCCCAATGCAACCGGACTTCCCGGTGTGTGCGGATCGGCGATGTGAC 240  
QY 706 ACCGGAACACGCGAGGTTCGGCGCGCGGTCTGCGGTCTGCGACCCAGCGGAC 765  
Db 241 ACCGCAACACCGCGGTTCGGCGCGCGCGTCTGCGGTCTGCGACCCACCAACCCAC 300  
QY 766 GAGCGTCCGCAAGTACGAGGAGTACCTTACTGTGTACGACGCGCGGGTCTCAG 825  
Db 301 GAGC-----GAAATTCAGGAACAGTACCTTTACTGGTCCACGATGGCGGTACTCC 354  
QY 826 TTCACGCGCTCCCGACCGCTCATCTGTAACACGCGGTGCGCGCGCCACACCGCC 885  
Db 355 TTCACGCGATTCCTGACCGGTCTTGTGAACACTGTGAGACGAGCGGCGACCCGCC 414  
QY 886 GCGGAGTCGAGACCGCGAGTGTTCGCGGACCCCAAGATCCACTGGGACACCGCGCC 945  
Db 415 GCGGAGTGGAGAACGAGAAGTGTTCGCGGACCGGAAATTCAGTGGGACGCGCGC 474  
QY 946 GGAATGGTCTGCTCATCGGACGAGTGGGTAGCGCGGTCTTACACTCTCGCGAAC 1005  
Db 475 AACGAGTGGTCTGTGTATCGGACGGCGCGCTACGTGCTTCTTACACTCTCTCCAAC 534  
QY 1006 CTGCGGACTGGACACTTTCGCGCAACTTCGACTACCGGAACACGCGCTCGCGGCA 1065  
Db 535 CTGCGGATGGCAATGGAGTCCAACTTCGACTACCCCAACACCGCTCGCGGTATC 594  
QY 1066 GAGTCCCGGAGCTGTTCGAGATCACCGAGACGAGGACACGCACTGGGTGTCTGCGC 1125  
Db 595 GAATCCCGGATCTGTTCGAAATGACCGAGGAGCGGAACCGGCACTGGGTGTTCGG 654  
QY 1126 GCCAGATGGAGCGCTACGGCATCGGCTCCCGCATGAGTACGCTACTGGACAGCGACC 1185  
Db 655 GCGAGATGGAGCGCTACGACATCGGCTTGGCCCATGACCTTTGCTTACTGGACAGTCA 714  
QY 1186 TGGGACGCGGAGCTTTCACCGCGACGACTTCACCGCGCAATGGCTTCAGTGGGCTGG 1245  
Db 715 TGGAGCGGACAGCATTCATCGCCGACAACTCACACACAGTGGCTTACTGGGATGG 774  
QY 1246 GACTGTACGGGCGGTCACTGGCCATCGATCGAGCGCGCGGACGACCAAGCGCTCGCC 1305  
Db 775 GACTGTACGGGCGGTCACTGGCGCGGTGGGAGACCTTGAGACCAAGCGGTTCGG 834  
QY 1306 ATCGGTGATGAACAACCTGGAAGTACCGCGCAGGCTCCCGACGACGACATCCGAC 1365  
Db 835 ACAGGTGATGAACAACCTGGAATATGCGCGCGCAACGCTGCCACGACCGCTCCGAT 894  
QY 1366 GGCTACAAACGGGAGAACTCGATCTCGCGGAGTGTGCGGTCTCGCGGACAGCTGCGCGC 1425  
Db 895 GGCTATACGGGCAAAATTCATCACGCGCGAGCTCAGGCTCGAGGCGCAATCGGCGCGC 954

```
Qy 1426 TGGTACACCTCTCGAGCACCCCGCTGGACGGCTGACGAACTAGTCAACCGCCACCACC 1485
Db 955 TGGTACACCTCTCGAGCACCCCGCTGGACGGCTTTCGAACTATGCCACCTCCAGCACC 1014
Qy 1486 ACACCTCCCGGACCGCTCGAGCGGAGCGCGCTCTCGCATCGAAGCGAGCGCATAC 1545
Db 1015 ACCCTTCGGGACCGCACAGTCAACGGCAGTTTCGCTACTTCGCTGAGGCGCGGCGTAT 1074
Qy 1546 GAGATCGAGCTCGACATCGCTCGGACACCGGACGACGCTCGGATCTCGGTGGCGCGC 1605
Db 1075 GAACCTGACCTCGATATTCATGGGACACGCGGACGAACTGGGAGTCTCGGTGGCGCGC 1134
Qy 1606 TCCCGCGAGGACCGCGCACACGAACTACGGAAGTACGAGGACGACCTCTAGCTCGAC 1665
Db 1135 TCGTCCGATGGAGCGCGCATACGAACATCGGCAATACGCTGACGAGTGTACGTCGAT 1194
Qy 1666 CGAGGACCTCGGACCTCGCGGCTACTCGTCCGCCCTCTACTCGGAGCGCGCGCCGCC 1725
Db 1195 CGCGCATCTCGGACGAAGCGTTATCGCTGGCACCTACACCGCGCGCGCGCGGCC 1254
Qy 1726 ATCGACCCCGCGCGGATCGCTGACCTCGGACGCTCGGATCTCGGACACGAGCGTCGAG 1785
Db 1255 ATCGATGCGAAGCGGATCGCTCGACCTCGGATCTTTAGACACCGCAAGTGTGAG 1314
Qy 1786 GTCCTGCTAAGCGCGGCGCACACGCTGCTCTCCAGGAGTCCACTTCCCGGAGGCGGAC 1845
Db 1315 GTGTTGCTAAATTCGCGGCGCACACGCTGCTTTCGAGCAGGCTGCACTTCGCGCGCGGAC 1374
Qy 1846 ACGGAATCTCGCTACACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1905
Db 1375 ACGGGATCTCGCTTATGCGGAGCGCGGCTCGCGCAACTTCACCGGATCACCATCGCG 1434
Qy 1906 GAGATTGGCGGAGCGGATC 1923
Db 1435 GAGTTCGGGAACCCATC 1452

RESULT 6
AB062062 1863 bp mRNA linear BCT 10-JUL-2002
LOCUS Microbacterium laevisformans mRNA for levanase, complete cds.
DEFINITION Microbacterium laevisformans mRNA for levanase, complete cds.
ACCESSION AB062062
VERSION AB062062.1 GI:14289196
KEYWORDS Microbacterium laevisformans (strain:ATCC 15953) cDNA to mRNA;
SOURCE clone_lib:pLEV-KB.
ORGANISM Microbacterium laevisformans
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Micrococcales; Microbacteriaceae; Microbacterium.
REFERENCE
1 Song,E.K., Kim,H., Sung,H.K. and Cha,J.
Cloning and characterization of a levanbiohydrolase from
Microbacterium laevisformans ATCC 15953
JOURNAL Gene 291 (1-2), 45-55 (2002)
MEDLINE 12095678
PUBMED 12095678
REFERENCE 2 (bases 1 to 1863)
AUTHORS Song,E. and Cha,J.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2001) Eunkyoung Song, College of Natural Science,
Pusan National University, Department of Microbiology; San 30
Jangjun-Dong Gumjung-Gu, Pusan 609735, Korea
(E-mail:eksong95@hanmail.net, Tel:82-51-510-3328)
COMMENT Submitted through BRIC(Biological Research Information Center) of
korea
URL: http://bric.postech.ac.kr
GeneNuri No. KS106615.
LOCATION/Qualifiers
1. 1863
/organism="Microbacterium laevisformans"
/strain="ATCC 15953"
/db_xref="taxon:36807"
```

```

/clone_lib="pLEV-KB"
/notes="re_sites:KpnI, SacI;vector:pBluescript;average
insert size:7931 bp-vector lab_host:E.coli"
1. 1863
/codon_start=1
/transl_table=11
/product="levanase"
/protein_id="BAB59060.1"
/db_xref="GI:14289197"
/translation="MGRPADGRRRPPGGCARTLTTTEGGSDAPALPARRALA
AGLALAGSAAALVAGPAAAEEDDAPSLRARYHTVDPHWNDRPQVVDGLHYLYL
YNADYDANPRANFGEETWRLATSDGVAFOGVAAKPTNANIDLWGSVVDHAGTA
GFGGAYMLVTQMDHTAAKLDASQQAQFLWYSVDGGRTERPDDEPVIQDGR
DFRDPKVVWDDERQWVLAERDVSFTSPDLRWRTREYVNAIGTIEPDLF
LRADGTHVMVGSANGYATNEPATYWTGSDGTFVDPVAAPOWLDHGFDMYGA
TWEDPAALERLIAIWMNNWDYAHGAPTPRGTGSDSIETREIARAGAGYSLL
SAPVDALQDHAHTHTALGDVROGFTELAYRGDAYEITTRISQDADNAGFOLRSAD
GSRHADAGLRDITAYLAKAQTGFRDSEKRSRTDADITVELRILVDRTTIEVFVG
GRYTHSSQVAFPSDQGLALYTGSPAFVRLRITFEADLAHVPRASSRSTRARGAR
GGRPPAPSPGVSRRRATCAGRRARAWPTSPGRTPRPASSRRRPSRSTGTRCTS"
BASE COUNT 270 a 728 c 627 g 238 t
ORIGIN
Query Match 18.5%; Score 370.4; DB 1; Length 1863;
Best Local Similarity 55.7%; Pred. No. 1.5e-31;
Matches 948; Conservative 0; Mismatches 691; Indels 63; Gaps 10;
Qy 301 CGCTGCCGGAACCGAGGCTGCTCGCGCGCCACCGCTCCACGAGAGACCAACCA 360
Db 34 CGCGCGCGGAGGCTGCGCGCGCCACCGCTGACCAACAGAGAGAGATCCCAT 93
Qy 361 ATGACGCGCGCAT---CTCAGCGCGCGCTGCTTCCAGGAGCGCGCGCGGACACTC 417
Db 94 GAGGCTCCCGCCCTGCCGCTCGCGCGCGCTCGCGCGCGGCTCGCCCTCGCCCTC 153
Qy 418 GCCTGATCTTCGGCGGTGCTGTCGCGCTCGAGCGCGGATCGGTCGCGGCTCGCTC 477
Db 154 GCGGCTCCGCGGCGCTGCTGCGCGGCGCGCGCGCGGAGGACGCGCGCGAGCTG 213
Qy 478 CGTGGCTGTACCATACGACCGCGCGCGCTGCTGCGACCGCGCGCGCGCGCGCTG 537
Db 214 CGCGCGCGCTATCATCTTCCAGCGCGCGCTGCTGAGTACGACCGCGCGCGCGCTC 273
Qy 538 ACCACCGCGCGCTTACAGCTGTACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTG 590
Db 274 GTGATCGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 333
Qy 591 -----CCCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
Db 334 CGTGCCTAACTTTCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTG 393
Qy 640 CACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 693
Db 394 GATCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
Qy 694 GCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 454 GCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
Qy 754 ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
Db 514 ACCAGATGGACCA-----CCCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
Qy 814 GCGGCTTACGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
Db 568 GCCAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
Qy 874 GCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
Db 628 CCCGTCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 681
Qy 934 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
Db 111 111 111 111111 111111 111111 111111 111111 111111 111111
```

CDS



Db 682 GAGACAGCGCGAGCGCTGGGTGCGCTTCATCGCGAGCGGACCGGGTTTGGTTCTAC 741  
QY 994 ACTCTCCGCAACTCGCGGCTGACCTGCGCGCACTTCGACTACCCGCAACACGCGC 1053  
Db 742 ACTCTCCGCGGCTGACCGCTGGAC-----GCGGACGGGGAGTACGTCAACGCGCGC 795  
QY 1054 CTCGGCGGCTAGAGTGGCCCCGACCTGTTGAGATACCGCGAGACGGGACAGCCAC 1113  
Db 796 ATCGCGCAGCATGAGTGGCGCGGACCTTCTCCGCTTCGGGCGGAGCAGCGGACGCCAC 855  
QY 1114 TGGGTGCTCGCGCGCAGATGAGCGCTAGCGATCGGCTCCCATGATGATGACGCTAC 1173  
Db 856 TGGGTATGGGCTGACGCGCAACGGGTACGCGAGCAACGACGCGGACCTACGCGGTAC 915  
QY 1174 TGGACAGGCACCTGGGACGGCGGAGAGTTCCACGGCGGAGCACTCACCCCGCAATGGGTC 1233  
Db 916 TGGACGGGCTGCTTCGACGGGCTGACCTTCGTTGCGGAGCTGCTGCGCGGAGTGGCTC 975  
QY 1234 GACTGGGGCTGGGACTGGTACGCGGCGGTACCTGATCGGCCATGATCGACGCGCGGAGACC 1293  
Db 976 GACCACGGGTTGACTGGTACGCGCGGTGACCTGGGAGGACCCCGCGCACCGCTCGAG 1035  
QY 1294 AAGCGCTCGCCATCGCTGGGATGACAACTGGAAGTACGCGCGGAGCGGACGTCGCCACC 1353  
Db 1036 CGCGCGCTCGCGATCGCTGGGATGAACTGGAATA---CGCGACGGGCGCGCGAC 1092  
QY 1354 GACGATCCGACGGCTACAAACGGGAGAACTCGATCGTCCGCGAGCTGGGCTCGCGCA 1413  
Db 1093 TGGCGCGGGAGGATTACCGCGGACCGGACTGATCACGGGAGATCGGCTCGCGCG- 1151  
QY 1414 GAGCTGGCGGCTGGTACACCTCTGTAGACACCCCGTGGGAGCGCTGACGAACACTGCTC 1473  
Db 1152 --TGCGGTTGCGGGTACAGCTCTCTCTCGCTCGCGTCCGCGGCTCCAGGACCGCG 1209  
QY 1474 ACCGCGACACACACTCCCGACCGGAGCTGAGGCGAGCGCGCTCTGCGCATGGAC 1533  
Db 1210 AGCGACACAGCGGCTCGGCGAGCTCGCGTGCAGCGGTTTACCGAGCTCGGCTACCGC 1269  
QY 1534 GGACGCGATACGAGATCGAGCTCGACATCGCTGGGACCGGACGCGGAGCTCGGCACTC 1593  
Db 1270 GCGAGCGGCTACGATACGACGCGGCACTTCGCGGACGAGGACGAGCAACGTCGGGTT 1329  
QY 1594 TCGGTGGCGGCTCCCGGACGGAAACCGGACAGCAATCGGCAAGTACGAGGAGAC 1653  
Db 1330 GAGCTCGCGGCTCGCGGAGCGGACGCGGCAACGCGGAGCGGCTCACCGCGCACTAC 1389  
QY 1654 CTGTAGCTGACGAGGACCTTCGACCTCGCGGCTGCTGCTGCGCCCTTACTCGGA 1713  
Db 1390 GCTTACTTACCGGCGGAGACCGGCGGCGGCGGCTGCTGGAAGG-----TCGAGA 1441  
QY 1714 GCGCGCGGCGGATCGACCGGCGGCGGCGGATCGCTGCACTCGGATCGCTCGGACAC 1773  
Db 1442 GCGGACGCGGCTGGA---CGACGCGGACACCGTGGAGCTCCGATCTCTGTCGACGCG 1497  
QY 1774 GAGAGGTGAGGTCTTGTAAACGCGGCGGACGCGTCTCCGAGGAGTCCACTTC 1833  
Db 1498 ACAGCATGATGAGGTCTTGTGCGGACGCGGCGCTACACGCACTCGAGCGAGGTGTCG 1557  
QY 1834 GCGGAGGCGGACAGGGAATCTGCTTACACCGGCGGCGGCGGCGGCTTACCGGCG 1893  
Db 1558 CCGTGGGGGACAGAGGCTCGCTTCTACACTTGGGCGGCGGCTGCGGTGTTCGCGAT 1617  
QY 1894 ATCTGCTGCGGAGATTGGCGAGCGATCTAGGGGATGACACACACGCTACCGGAA 1953  
Db 1618 CTGGGATCAACGAGTTGCGACAGCTCGCGGCGGCGGCTCTCTGCGGCGGCTCTCT 1677  
QY 1954 GCGGCGGCGGAGGACGAGG 1975  
Db 1678 AGGCGGACAGGTGGGCGCGGG 1699

RESULT 7  
AX211705

LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 1 from Patent WO0159126.  
ACCESSION AX211705  
VERSION AX211705.1 GI:15523937  
KEYWORDS Streptomyces noursei.  
SOURCE Streptomyces noursei.  
ORGANISM Streptomyces noursei  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 65140)  
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,  
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and  
Gulliksen,O.M.  
TITLE Gene cluster encoding a nystatin polyketide synthase and its  
manipulation and utility  
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;  
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN  
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;  
ALPHA PHARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich  
(NO) ; SEKUROVA, Olga Nikolayivna (NO) ; FJAERVIK, Espen (NO) ;  
BRAUTASET, Trygve (NO) ; STROM, Arne Reidar (NO) ; VALLA, Svein  
(NO)  
FEATURES  
source 1..65140  
/organism="Streptomyces noursei"  
/db\_xref="taxon:1971"  
/note="ATCC 11455"  
BASE COUNT 8270 a 25171 c 22273 g 9426 t  
ORIGIN  
Query Match 7.4%; Score 147.6; DB 6; Length 65140;  
Best Local Similarity 45.4%; Pred. NO. 5.7e-08;  
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;  
QY 8 ACCCGGACTTCCCTCGAGCACCGTCCCTACCGCGGCGGCGGCGGCGGCGGCGGCGG 67  
Db 19803 ACGGCGGAGACACCGACCGCGCGGCGGCTTCGGACTTCACCGCGGCGGCGGCGGCGG 19862  
QY 68 TCCTCAGCTAGACGGGCGGCGGCTCTCTGAGGTCTGCTCGGGGACGCTGAGGGGACTGCG 127  
Db 19863 AACAGCGCGCGCTACGCGGACCTCGGCGCATCAGCGCGGCGGCGGCGGCTTCGCGCT 19922  
QY 128 CGAAGCTGTCTCTTGGGGGCGGCGGCTGACCGGAGCTCGAGAGCGGACGCGGCGAG 187  
Db 19923 GGAAGAGGCTCTCGTTCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19982  
QY 188 GAACCGTGACGTGACCGGATCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 247  
Db 19983 CCGGCGGAGGACACCGTACCATCGCGGTACAGCGCGCGGCGGCGGCGGCGGCGGCG 20042  
QY 248 CTGCGCGGCTTCTGGGCTGACGAGCGCTCCACCGCGGACAGCTCTCTTCTACCGGTCG 307  
Db 20043 TCGACTCCCTGTCTCCCGCGAGGTCCCGCGCGGAGCACCGCGGCGGCGGCGGCGGCG 20102  
QY 308 CGAAGCAGGCTGAGCTTCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 367  
Db 20103 ACCGGGACTCCCTCTTCCACGTGAGTGGACCGCGGCTCCAGGGCGGCGGCGGCGGCG 20162  
QY 368 CGGCGCTCTACGCGCGGCGGCTTCCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 427  
Db 20163 CGGCGA-----CCGTCGCGGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20213  
QY 428 TCGGGGCTGCTGTGCGGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487  
Db 20214 TCCGCG-----CCACCGGATCCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCG 20264  
QY 488 ACCATGAGCGCGCGGCGGCTGCTGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 547  
Db 20265 CCGAGCGCGGAGGCGGCGGCTCCCGGACCTGGTGTGTACCAACCGGCGGCGGCGGCG 20321  
QY 548 GCGCGCTACAGCTGTACTACCTGCACTTCGAGCAACGCGGCGGCGGCGGCGGCGGCG 607  
Db 20322 GCGCGCGGCTCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 20381





```
/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
/translation="MASPDDLEERTAPRRRLVGLLRPHRRSVALVASMGVGGIVL
NAFGPLLLGRVTDIADGVLYGVPAGPAGIDFAALGRLLLVLLALVYVASLFLMAQGR
LVASAWRTIHELRRDAKRLRPLRPHRDOPAGELLSRTINDINLQOQLQOQLAE
LITISFLMLVLVSPSLVAVMLLSVPVSAIIAARIKRAQPHYAAQWMSANGTL
NAHVEVCTGHALLKGFORAAAEERFACNDVAVYRAAKAQAQFASGAPVMMFVANL
TYLLVAVGAKNPLNGTLTGDOVAFILYARQFSQPIVEIASVAGRLQSGTASQORVF
GTLDAPEAQNPPLRPGTPARAEGRVEFTDVSFRYSPDPLIENLSLTVPERGSTVAVI
PTGAKTTLGNLMLRFPEDPSGRILLDGTDTATRRDRRLRFGILQDTWLFQGTIA
ENIAYGAPGACARLDEEAARATCADRFRTIPLQGYDTVLDDESGTSGAGERKOLLTVAR
AFARPVAVLDEATSYDTRTEVLIQIRANSLRAGRTSFVIAHRLSTIRDAIDLIVM
DAGRIVECTHQDLQACQGLYARLHAARTHTPTAGAAAG"
complement(2662..4416)
/gene="nysg"
complement(2662..4416)
/gene="nysg"
/function="presumably involved in efflux of nystatin"
/notes="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAF71764.1"
/db_xref="GI:8050838"
/translation="MLRLRLRAQLPYAWATAALVALQVLQVILGLTLLPLTLGAALIDQ
GWNGDGRITELGVGMVAVLVOIAAALAAALAAATATAMGRDLRSALPRILDFS
AREIGRFTCSLLTRSVNDVOQVONLAQTGFIYVVCAPLCLGSLVLLALRODVLALL
IYALVLYAVCFGLLLARMTLYARMQTLDRGLRLREATIGTVRVYRVRDHERA
RPAQNTLVYVRRVGRGLIATMLPVLVLLMNGFTVALLTGTSHRIDAGRMPIGLSLA
ELCAYLTLMSVMLAFVFLVPRVRCAGRIAEVLDTCSSVAPPAAPQPVGRGAGRI
BLSYAGVPAEEVRLDVLTPVEGERIAVLGTSKGTLLNLVRLRLADATAGAV
RVGGTDRBELTAALAAVGPORPYLFSGTVASNLRFGRPDATDEBELWALVAQA
ADFAVMPDGLDAETTCGGNVSGORLRLSARALLRFRPEIYLFDDCFSDALDQATDA
ALRYALPYTAGATVITVAQRISAGRDADRIVLDGRVVAQGTGHDVLLRTPSYREI
ALSQTEEAAGLAGRP"
4714..5748
/gene="nysdiii"
4714..5748
/gene="nysdiii"
/function="presumably involved in mycosamine biosynthesis"
/notes="putative dcbp-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="Nysdiii"
/protein_id="AAF71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGQDGSYLABHLLSQGYQVWGLINGQANPKRSRV
SRLASELDFIDGDLMDQSLVSVDVTPQDEYVNLGALISFYPMQQAELVTEYNGMG
VLRMLEAIRMYSGLSLTSRTSPRGQIRFYQASSEMFGKAAETPQRETTLPHRSPYG
AAKAYGHYITRNYRESFQMYAVSGMLFNHESPRGQEPVTRKISLAVARIKQGLQDL
ALGNLDVARDNGYAGDYVRAMHLMLOQDAGDDYVIGTGMHSVRDARIAREHVLGNW
EDYVVDLPLVPAEVEVLCADSADQRLGWKPDVDFPTLIRMVWSDLAQVSRKNO
YGDVLLAANW"
5930..34363
/gene="nysi"
5930..34363
/gene="nysi"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/notes="polyketide synthase"
/codon_start=1
/transl_table=11
/product="Nysi"
/protein_id="AAF71766.1"
/db_xref="GI:8050840"
/translation="MDNEQKLRDYLKLATADLRTRRRRHVKLESAAEQPPVIAIMTCR
YPGVRSPEDLWRMVEAGEHCVTPPTDRGMDLEALAAAPASGGLFHDAPDFDADF
GISPREAVAMPQORVYLVESAWEAFERAGIDPTSVKSGRTGVEIGAMAQDYRVGPADG
AEGQLGTMGVSLSGRLSTFTFTGVPATVDTACSSSLVAVHLATQALRAGECTLAL
AGGVIWMSGPOTFIEMGRQGGLSADGRKCRSPFDATGVTGWAEGVGLILVLERLSDAVRN
```

```
CHETILAVVRGTAVNQDGASNGLTAPNGPSQOQVIOQALYNARLAAGDDIVVEAHGTGT
TLGDDVEAQAALLATYQGNRPADRPRLLLGVSXSNLSHTQAAAC?AGVIRKMWAMRHGTL
PRDTHAEPTTHVDMQSGAVRLTDDTDMPATPAPRAAVSVSGTSGTNAHMTIIEQAP
EPQPEDAATAQDDAGSTPATAPVVPVLLSGRTPD...3QAALARAALDTRGRP
DLDLAHSLATTRAGFEHRAVLLATLHPALDGLTALADADAPAAAPAWITGTTTRAET
RLAVLTSGQARLGCAGRLAARPAFATALDAALDAPTHLDRPLREVMGCTDAALL
DRYAYQAPALFAVEVALYRLIESQFVRPHDLAAGDQVGAIVAAHLAGVLSLADAATLVA
ARGUMOLAPDGGAMIAVOASEADVAPLLIAGHEDQVAIAVNGHSAVSVLSGAEATVTA
TVAQLAADRKTRLRVRSHAFSPHLMPLDAFRVAVDELTLQPLLVSNLTKPA
LAQIADRGTWVHVRAVRFADGIDMLARHDTTAFLELPGDGLSAMAQDCDLDAADA
DAVTLPALRACRPEHTLTALAGLHVHGTALDWTGCFAGTCARTDLPTVAFORRY
WPKALQSGTADLRVSGCAAHPLLSAAVSLADAGTLLTGLFSQTHPLMDHTRVG
TTLPGLTAFLELVRAGDEVCDRVEELTAAPLLPEQGGVQVQVLMGNPDVSRRRT
VNHARPOTDDTPTAHATGLVLTADASRLQPPASSEQGGT?LAGDPHPALDAQMP
AGAPFLPDHGYDLRDLGGFGYGPVQGLRMAWGGDVVYAEVLEPAGRSDAEAFGL
HPALLDAALHAAPTGLGERGGPFLFSEWEGVSLHAGGATLRVRLTVPVADALATV
ADGTGAPVSDVLSRVATQOLDTAAAVARDALFLDWTTPVOTPTDPGVALLGA
DPFGLLTHAGFADAPAYPDIAAALAAAGPVTTVVLSLACTGDDAADAARSARCAAE
ALAAVQTMWHHERFAARLVFTRGATVGRDVAAAAVWGLVRSQSNPCGFALVDL
DPDGAAGAAALVAAALVSGEPQLAVRDLVRVLRPLTEVAGADGTGDBGDGGSG
VSFSEGAALVTGGTGLGAVLARHLVAEYGVRLDLYVSRGSEAGAGELVAELAGV
GARVRVACDVTDRAAVVELVGHAVSAVHAAGVLDGDMVGALTGERLSAVLRPKVD
AVHLHEATRGDLDAFVFSGLGCVGPGQANYAANAFDLALMTRRABGLPGLS
LAWGPEQSGWGTGTLTDVDAERLARSQVPLSVAQGLALDAAVAGTDATCVPVRLD
LPVLRARGVEPPLRLSLIRVARRAAVAGSATAGNLAORLRLDEGDEWMLDVRG
QVALVLGHATGDDVDAGRAFRLDGLDFSALTVELNRMLTVTLGLRPLTLVDFDPTVRH
LATVLDLGLTDAEATVQAAVADDDPIVIGMACRYFGVSSPDELMRVLTEGT
DAVSGFTPNRGMWESLYHPDPHGTSTYRSGFLHEAGDFDPGFFGMSPREALATD
SOORLLPSSWEATERAGIDPVLSRSGRTYFAGVMYSDYAMLAISPEFEGFQSGSS
PSLASGRVATYLGEGPAAVTDATCSSLVAMHMAQALRSCECLLAGAGVYVNSTP
AVFDFARQRLSPDGRCKAFADAADGVGSEGVVLYLERQSDAVRNGHELAVVRG
AVFDDGASNLTPANGPSQQRVIRQALASGLTAGDVVVEAHGTGTGLDPTIEAQA
LLATYGRDEREPPLLLGVSXSNLGTQAAAGVGIKWLAMRHGVVPTRLHYDAPS
SHVMSSEGAALVLEQAAWPTETRVRRAGSVSGTGNVHVI?VEAQAKAIAAGA
ARRTPVAPVLLSGRSLARGLHGLQARDAELVDVLSLATTRSRFEORAA
VVAODRLOIASLGALADRPDPAVVEGEAGRTAVLFTGQSSORAMGRELHEVO
PEFAAADVACAVFDPDLDRPLREVVFADGSDDEAALLDETGWTPALFAVEVALFRL
PEFAGPDPFVAGHSIGETIAAHVAGVVLLEDAKCLAAARATLQKALTGGMATIAQA
TEDEIAHLDDTVAIAAVVQSVVSGDEAAETIAATFAERGGTKRLRVSHAFHS
PRMDGLDAFRIVAEGLTYRAPRIPLVSDLTGRADDAEVCYAEVYRVRHAEVAFAD
CVRTLRODAGATTFLELSDGLLTAAEMDTLGDHDAELVPMLRAGRAELAAATALAR
LQVGDVDAWAAVLAGTCARTDLPTVAFQHAYWYLPOLTPAALAAADAPDOOLWAA
VERGDARELADILGLGEQDLTPDLSLLPALT?SWRGNQEKHLIDTLYRVEVWTRLSKP
TAPVLDGTWLLVASDADATQDAPALLDGLDALGSHGARVRLRLDDSCADRVALERL
ADADVDAATQVLSVLPDERRDADCPPLTGLALTVALQALADTAGQGRUWLTATRG
AVSTNPADPTVTPQAAAGLGRGVLEHPRLWGLVLDPOVDFERAGQRLAGILAVK
DAPDGEDVALRATGVSQRLVTRVTEALPTAAEFTATGTLITGCTGLGAEARWL
ARAGQHLVLT?SRGPDPAGAAELRLEGVSPSVVACDVADRDLAAVLTALPEE
LPLTGVTHTAGVGHYGLDITLSTAEFAGLTAAGLAHLDALLADRELDFVFGSI
AGVWGSQNGAYGAANYLDALALHRRRAGLAATSVMGPWAEAGMAADDVASETLRR
OGLGLDPPAPMTELRRAVRODVTVTVDWQRYAPLFTSARPSALIAGLPEVRLAL
ADERTQDQATGASEVTVRRALAPEQLRLTDLVTPESATV.LGHSSADAVPEGRAF
RDVGFDSLTAVELRKRKGAATGLSLPTMVFDPYTPLEAOYLRAELIAGVLEAVGPV
ATGADDEPIAIIIGMACRFPGVSPESPEOLWIVASGTDASIEFFPVNRQWQTHLPD
PDRFTSTITOGGELHEADEFPITFGISPREALVMDPQKRLSLTETTESPERAGIRP
ETLRTSTITGTSYSSYQYGLGAGDTGEGHMTVGTSSVSLGRLLSVFGLGEPVATVD
TACSSSLVALHACQLRNGESNAVAGGATIMTTPNPIAFISQRLAKLGRCKRQAFS
DDADGMTLAEQGVVYVLRISDAQRNHPVLAIRGSAINDQNGSLTAPNGPSQQR
VIRQALANAPGDIDALEAHGTGTPGLDPLEAQALFATYGRDRPPESALLGVSVS
NIGHTQSAAGIASVIKWYMLRHSLEPPLHADAPSHVDSAGTCLVLLTOARAWPET
GRPRRAVSGFISGTSNAHVLEQAPVADTFAERPAVAPVPIAGVYVWVYVARSAA
ALRQAOERLLAHAEVETGALPAAGPLDGLSLVSARFAERHRAVVVPVPPAGDPLAALR
AVATDGPSPVAVGVADVGEHRTVFVPGQSGWGMGSQLDESDEFAERAECAAL
AEFTDMSLVDLVRGVGAPSLVERVDFVAFASVAGVSLAALWRSRGLVLPVAVHSGQ
EIAAAVYSGALSRLDGRVAVLRQAIGRALAGRGMMMSVALSVLDVLEPLVPEPVSQ
SVAAVNGRSPVVVAGEPEALDLARLTADDI RARRIADVASHSHOVEDLHEELLEV
LAEIAPRTSEVFFSTVTGMDLDTARMAGTFRNLRGRVDFADAVALLAAERAFV
EVSHPVLSMAVSAIDBAGVPAVAGTLRREDQGTDRFLLSAAEFVVRGVVDWAGL
PEGTGSARIDLPTVAFQHEHLWAVPPAPVAVAAADPDADFTAVTDEGDSVALLTAALG
TDEDSVAAVLPALT?SWRRARDRSTVDWARYRVAWKPLGGTLPHPLSTGTMLTADG
IDDDTDVAGALETGYAEVRLVDEECVDVAVRLRAGAEDVTGTVLSVLAARERTDAG
PGTSLVGLTALTVALIQALGDAAIDAPVWALTGRVASTGRADELTPVQAOVTGIGT
AALHEPQWGTGLDLPALADARAQRLLAALVLSGALGSDQLAIRPSGVFTTRIVRAEA
```



AUTHORS Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,  
Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and  
Gulliksen, O.M.

TITLE Gene cluster encoding a nystatin polyketide synthase and its  
manipulation and utility

JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;  
Norges Teknisk Forskningskapslige Universitet (NO); STIFTELSEN  
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);  
ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich  
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);  
Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Valla, Svein  
(NO)

FEATURES Location/Qualifiers  
source 1..125401  
/organism="Streptomyces noursei"  
/db\_xref="taxon:1971"  
/note="ATCC 11455"

BASE COUNT 15664 a 49692 c 42871 g 17174 t

ORIGIN

Query Match 7.4%; Score 147.6; DB 6; Length 125401;  
Best Local Similarity 45.4%; Pred. No. 4.6e-08;  
Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;

QY 8 ACCCGACTTCCCTCGAGGACACGCTCCCTTACCGCCGACCGCCGCGGCGGACTGCG 67  
DB 80064 ACAGCGGACACCGACCGCGGCTTCGGAGCTACACCGCGGCTGTGGACGCGGCAC 80123  
QY 68 TCCTCAGCTACAGCGGCGCTTCGAGGTCTTCGTGGGACGCTGAGGCGACTGGT 127  
DB 80124 AACAGCGCGCGCTTACGCGGACCTTCGGGCGCATCAGCGCGGCGGCTTCCGCTTCGGCT 80183  
QY 128 CGAAGCTTGGTCTCTTGGGGCGGCGGTGTGACGCGGAGCTTCGAGACGGACGCGGCGAG 187  
DB 80184 GGAAGGCGTCTCGCTCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80243  
QY 188 GAACGCTGACGTCAGCGGATTCGAGCTGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGG 247  
DB 80244 CCGGCGGAGGACACCTCACCATCGCGCTCTACGAGCGCGCGGCGGCGGCGGCGGCGG 80303  
QY 248 CTGCGCGCGTCTTGGGCTGACAGCGCTCCACCGCGAGCTCTCTTACCGCTGCC 307  
DB 80304 TCGACTCTCTTGGTCTTCCGCGAGGTCTCCGCGGAGCGACCGCGGCGGCGGCGGCGGCGG 80363  
QY 308 CGAAGCGGTTGAGCGCTTCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 367  
DB 80364 ACCGCGACTCTCTTCCAGCTGAGTGGACCGCGCTCCAGGCGCGCGGCGGCGGCGGCGGCG 80423  
QY 368 CGGCGATCTACGCGCGCGGCTGCTCAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 427  
DB 80424 CGGCGA-----CCGTCGCGCTCTCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 80474  
QY 428 TCGCGGCTGCTGCGCGCTGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487  
DB 80475 TCGCG-----CCACCGGATCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 80525  
QY 488 ACCCATGACGCGCGCGCGGCTGCTGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 547  
DB 80526 CCGAGCGGAGGCGCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80582  
QY 548 GGGCTTACAGCTGACTTACCTCGACTCCGACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 607  
DB 80583 GGGCGCGGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80642  
QY 608 AGCGGAGGACGACCGCGGCTTCAGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 667  
DB 80643 AGTGCTCGCGGACGCGCTTCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 80702  
QY 668 CCGACTTCCCGTGTGGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 727  
DB 80703 CCGAGGCGGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80762  
QY 728 GCGCGGCGGCGGCTGCTCGCGCTCGCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 787

DB 80763 AGAACCCCGGCGGCTTTCGCGCTTCTTCGACTCGCGCCCGACACCGCGCGCGCGCGCGG 80822  
QY 788 AGACGACTACCTTACTTGGTCGACCGACGCGGGGTTCACGTTTACCGCCCTCGCCCGACCGCG 847  
DB 80823 AGACCTTGGCCACCGCGCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 80879  
QY 848 TCATCTGTCAACACCGACGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 907  
DB 80880 GCACGAGCTGACGCGCGCGCGCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 80939  
QY 908 GGTTCGCGGACCCCAAGATTCACCTGGGACACCGCGCGGCGGAGATGGGTCTCGGTATCG 967  
DB 80940 GGAACCCCGGACGCGCTTCTGATCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 80999  
QY 968 GACGACTTCGGGTACCGCGGCTTCTACACCTTCGCGGACTCGCGGACTGGACACTTCGCGC 1027  
DB 81000 CCGCGGACCTTGTTCGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81059  
QY 1028 GCAACTTCGACTTACCCGAACACCGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1087  
DB 81060 CGGCGCGGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 81119  
QY 1088 TCACGCGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1147  
DB 81120 ACATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81179  
QY 1148 TCGGCGCTCCCGATGACTGACGCTTCTTGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1207  
DB 81180 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 81239  
QY 1208 CGGAGCGCTTACCCCGCAATGGCTCGACTTGGGGTGGGACTGGTACGCGGCGGCGGCGGCG 1267  
DB 81240 TCGGCTCTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81299  
QY 1268 GCGCATGATCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1327  
DB 81300 GGCACCTGCACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 81359  
QY 1328 AGTAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1387  
DB 81360 TCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81419  
QY 1388 TCGTCCGCGGAGTGGGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1445  
DB 81420 AGCGGCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81479  
QY 1446 CCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1505  
DB 81480 CGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81537  
QY 1506 GCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1565  
DB 81538 CGCGCGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81597  
QY 1566 TGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1622  
DB 81598 CCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81657  
QY 1623 GCACACGAAATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1682  
DB 81658 CCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81717  
QY 1683 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1736  
DB 81718 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 81777  
QY 1737 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1796  
DB 81778 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81837  
QY 1797 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1856



















QY 333 GCCACCCGCTCCACGAGAGAACACGATGACGCGCGCCATCTCAACGCGCGCGCTG-- 390  
Db 1131 GGTGCGGAACGGGCTACGGCCCGCTGCTCAACGTATGATCGGCGCGCGCTGGA 1190  
QY 391 ---CTCCAGGAGCGCGCGCGGAGCACTCGCCCTGATCTTCGGGGTCTGTCGCGCT 447  
Db 1191 CTTGCGGCGCGCGCGGCTCGGTGACACCGGCTGTCACCGGCGCGGTGGAGACTGNA 1250  
QY 448 GCAGCCCGGGATTCGCTCCCGGGTTCGCTCGTGGCGTTCACCATGACGCCGCCACG 507  
Db 1251 GTTCTCGGTGACGAGAGCGCGGCTCGCGCGCGGATCGACCTCGAGGCCAAGC 1310  
QY 508 GCGTGGCT-----CTGGACCCCAACGCGCGCGGTGACACCGCGCGGTACAGCT 560  
Db 1311 GCGCGGTGACGCGAGAGAGACGCGCCATCTGCGCGCTTCGTCGCTTCCTGGG 1370  
QY 561 GTACTACTGTGACTCGGACCAAGACGCGCGCGGTGCGGACACGCGAGCAGAC 620  
Db 1371 GCGCTTCGCGAGGCGCGCGGACACGCGCTCGCGCGGATCCAGCGCGCGCGCGCG 1430  
QY 621 CGAGCGGCTGCGCTTCACGACACACGCGACCGTGTGCGGTGCGCGCGCGCTTCCTCGGT 680  
Db 1431 CG 1490  
QY 681 GTGCTCGCGGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 740  
Db 1491 CG 1550  
QY 741 CGTCTGCGCTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800  
Db 1551 CGCGCGCTGCG 1610  
QY 801 CTGCTGCG 860  
Db 1611 CCGCGAAAGCGCTGCTGCGCGCTCAACCTGCGCGCGCTCGCGCGCGCGCGCGCG 1670  
QY 861 CGAGCGTCCG 919  
Db 1671 CGCGTCTCTAAATCG 1730  
QY 920 -----CCAGATCCACTGGGACACCGCGCGCGGAGATGGGTCTGCTCATCGGAGCT 974  
Db 1731 GCTGCGCTGGAGTCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1790  
QY 975 GCGGTACGCGCGGTTTACACTTCGCGGAACTGCGCGCGCGCGCGCGCGCGCGCGCG 1034  
Db 1791 CGAGGACCTCGCGCGCTACCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1850  
QY 1035 CGACTACCGGACCG 1094  
Db 1851 GCGCTACGCTATCTACACTTCGCGGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 1910  
QY 1095 AGACGACGCGGACCG 1154  
Db 1911 CAGCAAGTGTGCGCGCTTACCTTCCACCGACCGCGCGCGCGCGCGCGCGCGCG 1970  
QY 1155 CCGCATGAC---GTACGCGCTACTGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1211  
Db 1971 CGTGTGAGCGCTCTTCCACTCTACGCGCTTCGACTTCTCCGCTCGGAGATCTGGGCGC 2030  
QY 1212 CGACCTACCGCGCAATGGCTCG-----ACTGGGCTGGGACTGGTAGCGCGC 1259  
Db 2031 CCGTGTGACG 2090  
QY 1260 CCGTACCTGGCGCATGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1319  
Db 2091 CTTCTCGACCTGTGGCGCGGAGAGGTCAACCTCTCAACCGACACCGCGCGCTT 2150  
QY 1320 CAACCTGAAGTACCG 1379  
Db 2151 CCACAGCTGACG 2210  
QY 1380 GAACCTGATGTCGCGGAGCTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1439

Db 2211 CCGTCTGTTGCGCGGAGGCTCTCGACGTGCGCGCTCTCGCGGACTGGTAGCCGCGG 2270  
QY 1440 GAGCACCCCGCTGGAGCGCTGACGAACCTACACCGCGCGCGCGCGCGCGCGCGCG 1499  
Db 2271 CGGACGCGCGCGCGCGCTGGTCAACATGTACGCGATCACCGAGACCGCGTACGTCAC 2330  
QY 1500 GACCGTGCAGCGCGCGCGCTGCTGCGATGGAACGCGCGCGCGCGCGCGCGCGCG 1559  
Db 2331 GCACCGCGCGCTGCG 2390  
QY 1560 CATCGCTGGGACACG 1616  
Db 2391 CATCG 2450  
QY 1617 AACCG 1676  
Db 2451 CACCG 2510  
QY 1677 CGACCTCGCGCGGTACTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1736  
Db 2511 GCTGACCG 2570  
QY 1737 CCGCGGATCGCTGCG 1796  
Db 2571 CACCG 2630  
QY 1797 CCG 1856  
Db 2631 CGACGAGTCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2690  
QY 1857 GCTCTACCG 1916  
Db 2691 CACCG 2750  
QY 1917 GCGGAT 1922  
Db 2751 GCGGCT 2756

## RESULT 15

AF210249

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

77457 bp DNA linear BCT 29-AUG-2000  
Streptomyces verticillus bleomycin biosynthetic gene cluster,  
complete cds.

AF210249 AF149091

AF210249.1 GI:9937210

Streptomyces verticillus.

Streptomyces verticillus

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

Du, L. and Shen, B.

Identification and characterization of a type II peptidyl carrier

protein from the bleomycin producer Streptomyces verticillus ATCC

15003

Chem. Biol. 6 (8), 507-517 (1999)

99352421

10421758

2 (bases 1 to 77457)

Du, L., Chen, M., Sanchez, C. and Shen, B.

An oxidation domain in the BlmII non-ribosomal peptide synthetase

probably catalyzing thiazole formation in the biosynthesis of the

anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003

FEMS Microbiol. Lett. 189 (2), 171-175 (2000)

20389599

10930733

3 (bases 1 to 77457)

Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.

The biosynthetic gene cluster for the antitumor drug bleomycin from

Streptomyces verticillus ATCC15003 supporting functional

interactions between nonribosomal peptide synthetases and a

```
polyketide synthase
Chem. Biol. 7 (8), 623-642 (2000)
20500448
MEDLINE
PUBMED
11048953
REFERENCE
4 (bases 1 to 77457)
Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.
Direct Submission
Submitted (01-DEC-1999) Chemistry Department, University of
California at Davis, One Shields Avenue, Davis, CA 95616, USA
On Aug 29, 2000 this sequence version replaced gi:5326869.
COMMENT
Location/Qualifiers
source
1..77457
/organism="Streptomyces verticillius"
/strain="ATCC15003"
/db_xref="ATCC:15003"
/db_xref="taxon:29309"
complement(1199..2185)
/codon_start=1
/transl_table=11
/product="syrp-like protein"
/protein_id="AAG02342.1"
/db_xref="GI:9937211"
/translacion="MTCSTLTPLGGIIPRGEGLTTGAEXDGLGLDAGPDWVRAH
GPRLERLATDGLLHLGLPTDGDGVDGFDHVGSGDPLPTERSTPSRVKGNLY
TSEYFADQIPHNENSYAAHWPSTLYFFCHTAPDGTGAIPADGRAVLDLIPAEVR
RFSQGVYTRTFADMGLSWQAFQEDRGDVERHCRHQEQFSWDDGLRTPHRRP
ATAVDPTGAEVWFNOAHLPHPSLDPDLRQLVLETYGNGLPDRLADFADGCTIPDAD
LATVRAATRAALALPWREGDIMLVNDRMAHGREPTGERRVILVAMTSADS"
complement(2182..3918)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS12"
/protein_id="AAG02343.1"
/db_xref="GI:9937212"
/translacion="MTLTDLDAFLDQAARTPDAAHVHVGDTVMTYRELELRAGRMART
LAARGPGTLVAVRLPRGPEPVAALLAVVLITGAGYPLADDDPPDCRHLDDCAA
LLAEHPSRGRITLDEALAPARPAADVRAGDPAYVIYITSGSSGRHILVEQGA
LGAYLAQARAYDGLSRTVLHSLSDMAVTSLMGLVSGGAIHLDLKLAIASGTOP
PPAASRPSFLKVTPTSLHPLGLDPSCLPTGQVIGGEALTGSALGPWRAAHPDVTY
VNEGYPTAATGCAVTRPGDAVDPGAVIGRPFAGTGLYVLDADGEPVAVGVGEL
HIAGDOLARYLGRPLRTERFDPFADGSRMYRTGLVREPRDGLVLEGRADGQ
VKYSYIEPGEIENVLRHAGVRCACAVAGADARLVAYVVPDPSPGTAAPAR
HAAELPPYWPATFVTYVPELTPNGKLDRLALPGPPAGDAGPGDKPTPAETLCELL
ARALGPIEDADFLTSGTSTITALKLVAGARRVGIRLETLTVLRERTVRRILAAOP
DAASPLAEGYPE"
complement(3915..4733)
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG02344.1"
/db_xref="GI:9937213"
/translacion="MSKNAAHWSRIKRTGDAGPVVLAVDFTYGTQROEATFRHCLDLD
PVEWHAVPAPDGDWSTAGHRLRWTEGLDVTYLAGRPVRLVAGCVGFASALA
DALVERGHRPRVVLVPSAGPVATLDRFRLTAGMDLLTDGERRALLAETAIRRA
HAPDALVFAERVAALYREGCDLRCERLDVDSFGAEALAVLSLAYLTAALDVPT
PLMRGAVSLRSREHQGTDFDVEHGFDAVARAELLSSPVVAAALTALLREHEASR"
complement(4741..6024)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02345.1"
/db_xref="GI:9937214"
/translacion="MTARVATPTYLAVLAVREARAPLLGSLARMSFAYPLIALLLS
VRDTGSFVAGLTSGALSALTILFAPARALIDRGSRSLGVLTPVYLIGLAVLIT
LAREAPTALLIYAAVAGVPAPPLGPTMRVLRARILHGRQPLHHTAYALDSVTEVY
FTVGPLLGGIILAAAPLASMITVWVLIATAGTACFVLUSAATAAPAGEADEDRPHGR
PMALPGRTIIVLSFGVGLVGVVLPVLPADHAGSPGAGGILLSMLSGASVGGIA
YGRVLRSTPVRVNTVLTGFTLAVPLCLTASVPVAGAFALLVGLCLAPLFTAYLL
VNDLVASGTAPTTEANTVSTANNNGGFAAGSAAAGVLLDSRGPTVTVTAAFAVAAATA
VMTVLRRTLLLCAGHPPEAAATPADRTAAEAE"
complement(6021..6773)
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="AAG02346.1"
/db_xref="GI:9937215"
/translacion="MDPFLTRVMPFSGCDGRPRVRLCALPVAGGTAAYFKDWPAAL
PGVELLTAHLPGRGDRFTEPPATLEETAERLCEALPPSDLPVTVLHSGMALLGYE
VAARLAARGRAPNLLIAACRPPHPVDPDASGYTEAEADLAATLRARERWPTALRDEEVL
EAVLPALVADITAGDRYRPRRPLDPLKVIYIGADDDGTDRWTTLIGWRACATDCEV
VVIPLGGHYFLETDRAAVLTRVATDLAEAEVA"
complement(6784..7797)
/codon_start=1
/transl_table=11
/product="clavaminic acid synthase-like protein"
/protein_id="AAG02347.1"
/db_xref="GI:9937216"
/translacion="MTDLPLRTVALTGESAEVDDLRLTLADVPDSTVGLLHRTLIA
AOELPLIRAEILTCMRLYDSPRALVVTGFGVDDERIGTPTFAARPAPDPRTRDLELL
LLHAAALLGEAFGNATQNGRLVHDVLPVGEETAQMGSSSETELLWTEADAFPLRCD
YVGLLCRNHQRATVVGWPDLSRLTITTEDRAVLLLEPRYLIRPUSHPTPAONATGTRSA
ERFAATAEMDDAPERVAFLDPEDPYLRIDPAYMSPAGDAAARATDVTALIEDE
LRHVVLDAAGSLLLVNDNYQAVHGRKPFAAAYDGRDRLKRVNITRDLRRSRARSATS
LLV"
complement(7859..9214)
/codon_start=1
/transl_table=11
/product="putative transporter"
/protein_id="AAG02348.1"
/db_xref="GI:9937217"
/translacion="MMKSSRLRDLRGEDPVVAOESPQDAGTFCOGDDGLNVFAAL
AALLIEVPVVRPLPHAGLGRHVEHSSQAAAVALGPMQVASATITGVAGDGHQARRQ
RGVGVGREGVAGDDELGAEDRPHARQRLDGLWMAERLADLLVNPLOTVOGQDL
RGVSHDLGGDVLPGQRLSLGLSRRRGDVGTAHSAVGQPGPQGPAAAECSKG
LVARQDQDQALLRAVVEGPFQRGEDAGQVAETVDHPDVGHEVTVGGQOREVGGOL
GHWDRREVSVAGDGFQVAGVGLRLAPVSAAGHADVRAAGDGLLAVGREGQQQ
QRGGRAGDVHCPLVGVQVEDLTHGGODRLIVADLLRPQRTGLVDHRRPVMPLARV
DPGPDGPPSLAHSPLLTPTNSIPSTRGTPRCHLRKRPKRTSQSAARAPRRTGRELLVS
H"
complement(9211..17736)
/codon_start=1
/transl_table=11
/product="peptide synthetase NRPS11-10"
/protein_id="AAG02349.1"
/db_xref="GI:9937218"
/translacion="MASDALSALHLESDRCQRGGITVNDGSELAGONVAARFRERIS
AIPERTAILHKGATGYDELNRRAELTATLADAGAPSTLVAVALPRPDVLVATIC
ALLKGAACLPDPIGPAGRLREIMADAPDVLVTRAVAPFTGDPGLFLDDAPTIC
CSAVLPHRSAGTASEIAYLYPTTDEKSENSVVSVDMAVYLDPTAGTIPARAEILR
LVAPLLSGRLVLDADETRPVPTREAPDMDVEDVAVQVCAVLGVDRVGVDRDFDL
GKSLAAVQVVARLKLGLVLPRLALFDAPTVEELAAVRAEAGQGGQVREAALEP
VGRSEPLPSFAOORLWFLDRMLPDRAFYTMCDAFVRVGGIDGLARRRMLVGRHE
TLRTAFYERDGVPLYVGPADGFCARRVAAPTRVDSLJLHPEAREEAVRNILVAAEAT
PFRPADGALLRVVAVARLADDDHVLVSTHHIVSDAWSVGLVDELGRLYRECVTGDA
ALPPAVQYADFVAMQAMWAGPVQEEHAYWKRALDGPVSVI RLPMDHPRPAQSER
GETVGFALPDALVAALKLREGOATLFTLLGAFVOLLARHAGQEDIVVGVPAAGRT
RTETEPVGFVNTLPLRAICAGLSFERDLDLVQVREAAALGAFHQDPLFEALVEALAP
ERDUGHNPDLDRRMEYLVGHLYTLTGAAAADPGRTIGELPLSDGAEARLLDGFGR
GILNYGGTGFPERFEAVARTAPARATCGATTLTFAELNDRVERLAQALIGAGVTR
TPAVRLHPRSTDSVALLVAVRAGGVVPLDPDMPADRTAYILDVTAASVITRDLPA
QATFLHVDPRPADGLVAPRIDPQAAVYITSGTGAPKGVVVRHSILNHLTSL
LPLGHDHDPYLAGADGVPGDALKRTLTLTAPFTFDASMEQLSWMLACHELFIYEDVR
RDPSSALVRFVREHRIIDVITSSQLELLVSHGLDGEWAPSMVAVGEAVSPSLWRTL
RDQRTCFNLGYTEATVDATDCHLSDPADVPVIGTPLPHTHVLDRLRPVPPVGL
AGETYLGGTGLGARYLNRPALTARFVADVPDTPGSRVYLRVDRARPDGTLEYLG
RTDQOIKIRGFVRPEGEIEALVTHHPAVKEAAVDDAHARLVAVVTLAEGGGAGPTDV
RRAQGRPLPAHMPVSAVLEALPLTSNGKLDRLAPAPAGRELVDVREVPADMEV
EVVAQWCAVIAETGLGGAAPPLGSDRSGPLPLSQAQQLWYLDQAPDSVSNMCD
BEAARVRLDIALRLARTVERETLRTAFVERDGVPHQVVSAPDAPARAEEVVR
IYAVRGYGLDIALRLARTVERETLRTAFVERDGVPHQVVSAPDAPARAEEVVR
IEAQRTEADEROLVAARTPRPADGLMARVARLADDDHVLVVTTHHIVSDGVS
VDLILDELGRLYREHVTGDPAGLPDVOYADFAVMORSMTGTPVREHLYAKNRALD
GAVSLRLPADHPRPAVQSGRGETVEFPLPAPVARLEALCRGGQVTLFWALLGAFQV
LLARYSGQDDVVVGVPTANTRAETEPELVGFVNTLPVAVACSPESLFWALLGAFQV
```

ALGAFADHDLPEALVEALAPERDLGHPLVQVTFQLLDADPERLVLHGTCVSLGFG  
GVTFSLDLSVSGRGKRCVLTYPDLDFPRMEVLAGHYLLTLAGAADPLGLRVG  
DLPSRDDVRLLLGSGSPRYLPAPAEVTPDAFAAQVRAFPALVHGSDTLTFAE  
LDRTVLTAAALRCGVAAETPVAVCLPRSDADVALLAVLRAGVVPVDPENPSGR  
VAHVLDTEAAVPIVTRDLPADPGRVHLPQAPADDPRLPLRLHQAYLIIFTSGST  
GAPGVVVRHGLSHLGHVRMAEGGRNVATMTDFPSLEQFLMVLVAGHTLHV  
APEEVRDRPEALVALVRAAIDLNVTPSHLTLTEAGLEEDRPGVTLVGGRAVPA  
ALMRTLREBTGATRFNLYGTEATVDATSHLTLDPADVPVIGTLPHTHVRLDRL  
RPVGVAGELYLGCTGLARGYLNRPALTQREVADVPDTPGSRVLRGRARWRPD  
GTLEVLTGTDQIKIRGERVEPGETEAVLTHHPAKVKAATVANDDGAARLAVVPA  
PRAPGDSADAPAOQEVENNAVFEATHDADGELIFNIKWNDSLTGAPIPAEHR  
EMWDTTVALRLFAERVLIEGSGTGLMRLMLPHVETITGTFDSRPVDMWLRRLR  
RPAHRLTLREATHDEFTVRAASTDLVVVNSVQYFPDRAYLDTVLARALDADRGR  
VFVGDRLMALQFYARQALAHAGPAAARVADRAAGFAAFMDELILVSPAYFAALA  
ARSPIRVTEILPRKRNEMSLRYDVLVHVGDRPAEPAEVLWTGVDHDLASL  
SARLGRGPDALLVRGVANDRLTRDNELLDPARTAVEPEDLGLADSLTPYRVSVW  
AAADPRGMVLLVRRDHDGDLPLVPHVPPEPSAPLINTPTRIHPSARQSGSADGLR  
SWAERLPAHLUPARITEVDALPRTGKLPORGALGLVITAGRGARAGDRPATRTG  
LERTLADAWARVLGLPEVGVHENFFALGGDSLLAVRAVRCRAGVRLTVROLLSEQT  
VAALAAALEEESQ\*  
complement(18307..18660)  
/codon\_start=1  
/transl\_table=11

Query Match 5.6%; Score 112; DB 1; Length 77457;

Best Local Similarity 42.4%; Pred. No. 0.00038;

Matches 794; Conservative 0; Mismatches 1065; Indels 13; Gaps 3;

QY 9 CCCCAGCTTCCCTCGACGACACACGCTCCCTACCGCGGACCGCGCGCCGACTGCT 68  
DB 38086 CCCCAGGACGGCGCTGCACAACTGCTCCACCTCGCCCGGGGCTTCGGCGCGTGA 38145  
QY 69 CCTCAGCTAGACGGCGCCCTCTCGAGGTCTTCGTGGGGACGGTGGAGCGACTCGCTC 128  
DB 38146 GGAGCGCACCCCGCGCGTGTGACCGTGACACCGGTGCCACGAGTCTCGGGA 38205  
QY 129 GAACCTGTCTCTGGGGCGCGGGGTGTGACCGGAGCCTCGAGACGGACGGCAGG 188  
DB 38206 CGACCTCGCCACCCCGAGCAGCACCGTCCCGCGCGGCAAGTGTATCCCGCGGA 38265  
QY 189 AACCGTGACAGTGACCGGATGACGTGAGGCGCCGACGATGCTGAGCGCCCTGAAC 248  
DB 38266 GTACCGGTGGATCGCTTGACCGCTTGACGTGAGCGGGGCTTGAGCGCGGCTT 38325  
QY 249 TCCCGCGTCTTGGGTGACGAGCGCTCCACCCCGACAGCTCTCTCTTACCGTGGCC 308  
DB 38326 GCGGACCTGATCGTCCGGGAACCTCGGCGCGCGCGAGACACCGTCAACCGCTGCG 38385  
QY 309 GAACAGGTTGACGCTTGTCTGCGGCCACCGCTCCACGAGAGAACAGCAATGACGCC 368  
DB 38386 CGGCGGACCGCTTACACCCCTGCGCGTCCGGAGCGCCCTCCCGCGCGACCGAAG 38445  
QY 369 GGCATCTCACCGCGCGCTCTCAGGAGCGCGCGCGGACGACATCGCCCTGATCTT 428  
DB 38446 CCGCGGCTCGCGCGCGCGCTTACCTGTCTGCGCGGCTCGCGCGCATCGGCT 38505  
QY 429 CGGCGGTGCTGTCGCGCTGACGCGCGGCGATCCGCTCGGGCTCGCTCGCGCTCTA 488  
DB 38506 CCACCTCGCGAGTAC---TGGCGCGCGCGCGCACACCGCTCGCTCAACCGCGCG 38562  
QY 489 CCACATGAGCCCGCGCGGCTGCTGCGACCCCGCGCGCGCGCTCACACCGCG 548  
DB 38563 GCGCTTCCCGCGCGCGGTGGAGGGCTGCGCGGACACCGCGAGCGCGCT 38622  
QY 549 CGCCTACAGCTGTACTACTCGACTCGGACCAAGACGCGCGCGCGCTGGGACCA 608  
DB 38623 GTTCGGGGTTCGCTCCCTCGCGCGCACCGCGCGCGCTGCTGCTCCCGCGCGCA 38682  
QY 609 CGCGAGCAGACGCGGTGCGCTTACGACACCGACCGCTGATCGCGCTCGCGC 668  
DB 38683 CCTCACGACACGAGCGATGCGCGCGCTTCGCGGACGAGGTGAACAGCGCGCGC 38742  
QY 669 CGACTTCCCGTGTGGTTCGGGTTCGGCGGTCTCGCGCACCGCGAGCGGTTGCG 728

DB 38743 COTCCGGGGGGTGGTGACACCGCGCGGGTGCCTCCGACACCGCGGATATCCAGCGTCG 38802  
QY 729 CCGCGCGCGGGTCTGCGCGCTCGGACCCAGCGACCGCGCTCGCAAGTACCAGGA 788  
DB 38803 CCACGACCGCGGACGAGCGCGCGCTCCCGCGCAAACTGACCGGACCGCTGCTCTGGA 38862  
QY 789 GCAGTACCTTCTACTGTGTGACGAGCGGGGGTTCAGTTTCAACCGCGCTCCCGACCGCT 848  
DB 38863 CGAGGTGTTCGCCACCGCGGACCTCGACTTCTCGTCTGTCTCTCGATCGGACCGT 38922  
QY 849 CATCGTAAACACGAGGTGCGCGCGCCACACAGCGCGCGCGAGATCGAGAACCGCGATG 908  
DB 38923 GCTGCACAAGCTCAAGTTTCGGGAGGTTCGGTACGTGCGGGGCAAGAGTTCCTCAGCG 38982  
QY 909 GTTCGCGGACCCCAAGATCCACTTGGGACACCGCGCGGAGATGGTCTGCTCATCGG 968  
DB 38983 CTATCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGAGAACCTGTCTGATCGCTGGACCGA 39042  
QY 969 ACAGTTCGCGGTACGCGCGCTTCTACACTTCGCGGAACTTCGCGGACTTGCAGATTCGCGG 1028  
DB 39043 CTGCGGGAGTTCGGGATGTGGCGCGCGCGCGCGCGCGCGCTGTGACCGAGCGCTACGCGAC 39102  
QY 1029 GAACCTGACATCCCGAACACCGCGCTCGCGGCGATCGAGTGCCTGCTGCTGAT 1088  
DB 39103 CGCGCGCGGACCTGCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGATCAGCGCGCGGAGGA 39162  
QY 1089 CACCGCAGACGAGGACACGCACTTGGTGTCTGCGCGCGAGCATGGAGCGCTTACGGCAT 1148  
DB 39163 GGGCGTGCAGCTTTCGCGCGGCTGCTCGCGCGGACACCGCGCGAGCTCATCGTCTC 39222  
QY 1149 CGGCTCCCGATGAGCTAGCGCTACTTGGAGAGACACTTGGGACGGGAGAGTTCACAGC 1208  
DB 39223 GCGCGAGGACCTCGAGAACTCTCGCGCGGACGCGCGGTACACCGAGGACCACT 39282  
QY 1209 CGACGACCTCACCGCGCATGCTGACTGGGTGGGTGGGACTGGTAGCGCGCGCTCACCTG 1268  
DB 39283 CCGCGCGCTTCGGGACCTGAGGATCGCGCGCGGGGACCGCTCGCGCGCGCGCGCGC 39342  
QY 1269 GGCATCTGATCGAGCGCGCGGAGACCAAGCGCTTCGCGCATCGGTGGATGAACATGGA 1328  
DB 39343 GTACGGCGCGCGCGCACCGCGCGCGCGCGCGGATCGCGCGCTGTACCGGACCTGCT 39402  
QY 1329 GTACCGCGCGCGAGCTCCCCACCGACGATCGAGCGGTACAAGGGGAGAACTCGAT 1388  
DB 39403 CGCGCTCGAAACAGCTCGGCTCGAGGACGACTTCTTCGGGCTCGGGGAGCTCGCTGCT 39462  
QY 1389 CGTCCGCGAGCTGCGGCTCGCGCGGACGCTGCGGCTGTACACCTCTCTGAGCACCGC 1448  
DB 39463 CCGCTTGGCGCTGTCTGCGAGTGGGAGCGCTACGGGTGGAGATCTCCCTCGCGCG 39522  
QY 1449 CGTGGCAGCGCTGAGGAATAGCTCACCGCGCACACACACTCCCGGACCGGACCGTGA 1508  
DB 39523 CATGTTGACGAGGCCA-----CGTGGCGGGCTGGCGCGCGCACCGCGCGCGC 39575  
QY 1509 CGGAGCGCGCTCTGCGCATGGAAGGAGCGGATACGAGATCGAGCTCGATCGCTG 1568  
DB 39576 CGGAAGAGAGCGCGCGCGAGAGAGTGTGT---GTGACCGCGCGCGCGCATACCGA 39632  
QY 1569 GGACACCGGAGAGAGCTCGGATCTCGGTGGCGCGCTCCCGGAGGAAACCGGACAC 1628  
DB 39633 CTTGCTACCGAGCTTCGCGCGCGCGGAGTGAACCTACCGCGGAGCGGAGCGGCTGCA 39692  
QY 1629 GAACATCGSAAAGTACGGAGCAGACTGTACGTTCGACCGAGGAGCGCTTCGACCTCGCGCG 1688  
DB 39693 CTGCGCGCGCGCGCGGCGCTCACCGAGAGTCTCTCGCCACCATCCGCGCGCGCG 39752  
QY 1689 GTATCTGCTGCGCGCTTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCT 1748  
DB 39753 CGACGAACTCTCTCGCGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39812  
QY 1749 GCACCTTGGCATCTCTGTCGACACAGAGCTGAGGCTTCTGTCACCGCGCGCGCGCAC 1808

Db 39813 GCCGCTGCTTCGCCCAAGGAAGGCTCTGGCTCCTCCACCAAGTTCCACCCGCGACGACAG 39872  
Qy 1809 CGTGCTCTCCCAAGCAGGTCCACTTCGCCGAGGGGCGACACAGGGAATCTCGCTCTACACCGA 1868  
Db 39873 CGCCTACAACATCCCCCTGCACATCGCCCTGCGGGGCCCTGAACCGCGCCGCTGCG 39932  
Qy 1869 CGGCGGCCCCGC 1880  
Db 39933 CGCGCCCTGGC 39944

Search completed: May 10, 2003, 08:41:16  
Job time : 6017.5 secs



**THIS PAGE BLANK (USPTO)**



PA (REAL-) REALBIOTECH LTD.

XX Rhee S, Song K, Kim C, Ryu E, Lee Y;  
XX WPI: 2001-308483/32.  
DR P-PSDB; AAB82301.  
XX Producing difructose dianhydride IV from sucrose, involves reacting  
PT sugar solution in the presence of levansucrase to produce levan, and  
PT reacting levan solution in the presence of levan fructotransferase to  
PT produce DFA IV -  
XX  
PS Claim 4; Page 48; 72pp; English.  
XX  
CC The present sequence is that of *Arthrobacter ureafaciens* K2032 DNA  
CC encoding levan fructotransferase (see AAB82301). The DNA was  
CC obtained by PCR amplification of genomic DNA using degenerate  
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying  
CC the levan fructotransferase gene is claimed. Also claimed is  
CC *Escherichia coli* JUD81 (K12C 0877BP), prepared by transforming  
CC *E. coli* DH5-alpha with pUDFA81. A claimed process for producing  
CC difructose dianhydride IV (DFA IV) from sucrose comprises  
CC subjecting sugar solution to reaction at room temperature or lower  
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase  
CC derived from *Zymomonas mobilis* to produce levan, purifying the  
CC levan from the reaction solution, and subjecting it to reaction at  
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in  
CC the presence of levan fructotransferase, preferably obtained by  
CC cultivation of *E. coli* JUD81. The product is useful as a  
CC low-calorie sweetener.  
XX  
SQ Sequence 2001 BP; 337 A; 813 C; 588 G; 263 T; 0 other;

Query Match 100.0%; Score 2001; DB 22; Length 2001;  
Best Local Similarity 100.0%; Pred. No. 2e-310;  
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTGCACCCGACCTCCCTCGACGACACCCGTCCTACCGGCGGACCGCCCGGCC 60  
DB 1 GGGTGCACCCGACCTCCCTCGACGACACCCGTCCTACCGGCGGACCGCCCGGCC 60  
QY 61 CGACTGCTCCTCAGCCTAGACGGGCGCCCTCTCAGAGGTCTTCGTCGGGACGGTGAGCG 120  
DB 61 CGACTGCTCCTCAGCCTAGACGGGCGCCCTCTCAGAGGTCTTCGTCGGGACGGTGAGCG 120  
QY 121 ACTGCGTCGAACCTGGTCTCTCTGGGGCGGCGGTGTGACCGGAGGCTCGAGACGCA 180  
DB 121 ACTGCGTCGAACCTGGTCTCTCTGGGGCGGCGGTGTGACCGGAGGCTCGAGACGCA 180  
QY 181 CGGCCAGGACCGGTGACGTGACCGGATGACGTGAGGCGCCAGCGATGCTGACGCC 240  
DB 181 CGGCCAGGACCGGTGACGTGACCGGATGACGTGAGGCGCCAGCGATGCTGACGCC 240  
QY 241 CCTGAACCTGCGCGCTTCTGGGTGACGAGCGCTCCACCGCCGACAGCTCTCTTCTAC 300  
DB 241 CCTGAACCTGCGCGCTTCTGGGTGACGAGCGCTCCACCGCCGACAGCTCTCTTCTAC 300  
QY 301 CGCTGCCGGAACAGGGTGGAGCTTCGTGCGGCCACCCGCTCCACGAGAGAACGCA 360  
DB 301 CGCTGCCGGAACAGGGTGGAGCTTCGTGCGGCCACCCGCTCCACGAGAGAACGCA 360  
QY 361 ATGACCGCGGCGATCTACAGCGGCGCGCTGTCTCAGGAGGCGGCGCGGAGCACTCGCC 420  
DB 361 ATGACCGCGGCGATCTACAGCGGCGCGCTGTCTCAGGAGGCGGCGCGGAGCACTCGCC 420  
QY 421 CTGATCTTCGGCGGTGTCGCGCTCAGCGCGGATCGGCTCCGGGTCTCGTCTCGT 480  
DB 421 CTGATCTTCGGCGGTGTCGCGCTCAGCGCGGATCGGCTCCGGGTCTCGTCTCGT 480  
QY 481 GCGCTTACCAATGACGCCCGCCAGCGGCTGGTCTCGGACCCCGCCAGCGCGGTACC 540  
DB 481 GCGCTTACCAATGACGCCCGCCAGCGGCTGGTCTCGGACCCCGCCAGCGCGGTACC 540  
QY 541 ACCACGGCGCTACCACTGCTACTGCTCCGACGACGAGAACGCGCCCGGCCG 600

DB 541 ACCACGGCGCTACCACTGCTACTGCTCCGACGAGAACGCGCCCGGCCG 600  
QY 601 TGGGACACGCGAGACGACCGCGCTCGCTTTCAGCGACACGCGACCGTGTATGCGG 660  
DB 601 TGGGACACGCGAGACGACCGCGCTCGCTTTCAGCGACACGCGACCGTGTATGCGG 660  
QY 661 CTGCGGCGCGACTTCCCGGTGTGTCGGGTGCGGCGGTGTCGGGACCGCGAGAACGCA 720  
DB 661 CTGCGGCGCGACTTCCCGGTGTGTCGGGTGCGGCGGTGTCGGGACCGCGAGAACGCA 720  
QY 721 GGGTTCGGCGCGCGGTGTCGCGCTCGGACCGACGCGCGCGGTCTCGCAAG 780  
DB 721 GGGTTCGGCGCGCGGTGTCGCGCTCGGACCGACGCGCGCGGTCTCGCAAG 780  
QY 781 TACAGGAGCAGTACTCTACTGTCGACCGACGCGCGGTTCACGTTTACCCGCTCGCC 840  
DB 781 TACAGGAGCAGTACTCTACTGTCGACCGACGCGCGGTTCACGTTTACCCGCTCGCC 840  
QY 841 GACCCGCTCATGTCGTCACACCGACCGGTGCGCGCGCACACGCGCGGAGATCGAGAAC 900  
DB 841 GACCCGCTCATGTCGTCACACCGACCGGTGCGCGCGCACACGCGCGGAGATCGAGAAC 900  
QY 901 GCGGAGTGTTCGCGACCGCCCAAGATCCACTGGGACACGCGCGCGGAGATGGTCTGC 960  
DB 901 GCGGAGTGTTCGCGACCGCCCAAGATCCACTGGGACACGCGCGCGGAGATGGTCTGC 960  
QY 961 GTCTCGGACGACTGCGGTGACGCGCTTCTACACTCGCGGAACTGCGGACATGGACA 1020  
DB 961 GTCTCGGACGACTGCGGTGACGCGCTTCTACACTCGCGGAACTGCGGACATGGACA 1020  
QY 1021 CTTCCGCGCACTTCGACTACCGGAAACACGCGCTTCGCGGCGATCGAGTGCACCGCTG 1080  
DB 1021 CTTCCGCGCACTTCGACTACCGGAAACACGCGCTTCGCGGCGATCGAGTGCACCGCTG 1080  
QY 1081 TTCGAGATCACGCGACGACGGGACCGCACTGGGTGCTCGCGCGCAGCATGACGCGC 1140  
DB 1081 TTCGAGATCACGCGACGACGGGACCGCACTGGGTGCTCGCGCGCAGCATGACGCGC 1140  
QY 1141 TACGCGATCGGCTCCCTCATGACGTACGCTTACTGAGACGCGACCTGGGACGGCGAGC 1200  
DB 1141 TACGCGATCGGCTCCCTCATGACGTACGCTTACTGAGACGCGACCTGGGACGGCGAGC 1200  
QY 1201 TTCCACGCGGACGACTCACCGCGCAATGCTGCTGACTGGGGCTGGGACTGGTACGCGGC 1260  
DB 1201 TTCCACGCGGACGACTCACCGCGCAATGCTGCTGACTGGGGCTGGGACTGGTACGCGGC 1260  
QY 1261 GTCACCTGGCCATCGATCGAGCGCGCGAGACCAAGCGCTCGCCATCGCGTGGATGAAC 1320  
DB 1261 GTCACCTGGCCATCGATCGAGCGCGCGAGACCAAGCGCTCGCCATCGCGTGGATGAAC 1320  
QY 1321 AACTGGAAGTACGCGCGACGCGACGTCCCGACCGACGATCCGAGCGGTACACCGGCGAG 1380  
DB 1321 AACTGGAAGTACGCGCGACGCGACGTCCCGACCGACGATCCGAGCGGTACACCGGCGAG 1380  
QY 1381 AACTGATGTCGCGGAGCTGGGGTTCGCGGACACGCTGGGGCTGGTACACCTCTCTG 1440  
DB 1381 AACTGATGTCGCGGAGCTGGGGTTCGCGGACACGCTGGGGCTGGTACACCTCTCTG 1440  
QY 1441 AGCACCCGCTGGGACGCTGACGAACTAGCTACCGCGCACCCACACTCCCGGACCGG 1500  
DB 1441 AGCACCCGCTGGGACGCTGACGAACTAGCTACCGCGCACCCACACTCCCGGACCGG 1500  
QY 1501 ACCGTCGACGCGCGCGCTGCTGCTATGGAACGCGACGCGATACGAGATCGAGTCTGAC 1560  
DB 1501 ACCGTCGACGCGCGCGCTGCTGCTATGGAACGCGACGCGATACGAGATCGAGTCTGAC 1560  
QY 1561 ATCGGCTGGGACACGCGGACGAACTCGGCTGCGGCGCTCCCGGACGCGAAC 1620  
DB 1561 ATCGGCTGGGACACGCGGACGAACTCGGCTGCGGCGCTCCCGGACGCGAAC 1620  
QY 1621 CGGACACGAACTCGGCAAGTACGAGGACGACCTGACGCGACGAGGACCTCCGAC 1680

Db 1621 CGGCACGAAACATCGGCAAGTACGAGCAGACACCTGTACGTGACCGGAGGACCTCCGAC 1680  
Qy 1681 CTCGCCGGGTACTCGTCGCCGCCCTACTTCGGAGCGCGCCGCCCATCGACCCGGGCGCC 1740  
Db 1681 CTCGCCGGGTACTCGTCGCCGCCCTACTTCGGAGCGCGCCGCCCATCGACCCGGGCGCC 1740  
Qy 1741 CGATCCGTCACCTCGCGCATCTCGTCGACACCCAGAGCGTCGAGTCTTCGTCGAAGCC 1800  
Db 1741 CGATCCGTCACCTCGCGCATCTCGTCGACACCCAGAGCGTCGAGTCTTCGTCGAAGCC 1800  
Qy 1801 GGCACACCGTCTCCACGAGTCCACTTCGCGAGGCGGACACGGGAATCTCGCTC 1860  
Db 1801 GGCACACCGTCTCCACGAGTCCACTTCGCGAGGCGGACACGGGAATCTCGCTC 1860  
Qy 1861 TACACGACGCGCGCGCCGACACTTCACCGGCATCGTCGCGGAGATTGGCCAGGCG 1920  
Db 1861 TACACGAGCGCGCGCCGACACTTCACCGGCATCGTCGCGGAGATTGGCCAGGCG 1920  
Qy 1921 ATCTAGGCGATGACACACACCTCTACCGAAGCGCGCCCGGGAGACGACGCGCGAC 1980  
Db 1921 ATCTAGGCGATGACACACACCTCTACCGAAGCGCGCCCGGGAGACGACGCGCGAC 1980  
Qy 1981 AATCGACACGTCCTCGTGGTT 2001  
Db 1981 AATCGACACGTCCTCGTGGTT 2001  
RESULT 2  
AAF30919  
ID AAF30919 standard; DNA; 2000 BP.  
XX AC AAF30919;  
XX XX  
XX 09-JUL-2001 (first entry)  
XX Arthrobacter ureafaciens levan fructotransferase DNA.  
XX Levan fructotransferase; difructose dianhydride IV; sweetener; ds.  
XX OS Arthrobacter ureafaciens.  
XX XX  
XX Key Location/Qualifiers  
XX CDS 360..1925  
XX sig\_peptide /\*tag= a  
XX mat\_peptide /\*tag= b  
XX /\*tag= b  
XX W0200129185-Al.  
XX XX  
XX 26-APR-2001.  
XX 19-OCT-2000; 2000WO-KR01183.  
XX 19-OCT-1999; 99KR-0045302.  
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.  
XX (REAL-) REALBIOTECH LTD.  
XX Rhee S, Song K, Kim C, Ryu E, Lee Y;  
XX WPI; 2001-308483/32.  
XX Producing difructose dianhydride IV from sucrose, involves reacting  
XX sugar solution in the presence of levansucrase to produce levan, and  
XX reacting levan solution in the presence of levan fructotransferase to  
XX produce DFA IV -  
XX XX  
XX Disclosure; Page 8-9; 72pp; English.  
XX PS  
XX The present sequence is that of Arthrobacter ureafaciens K2032 DNA  
XX encoding levan fructotransferase (see AAB82301). The DNA was

CC obtained by PCR amplification of genomic DNA using degenerate  
CC primers (see AAF30921-22). Expression vector pUDFA81 carrying  
CC the levan fructotransferase gene is claimed. Also claimed is  
CC Escherichia coli JUD81 (KCTC 0877BP), prepared by transforming  
CC E. coli DH5-alpha with pUDFA81. A claimed process for producing  
CC difructose dianhydride IV (DFA IV) from sucrose comprises  
CC subjecting sugar solution to reaction at room temperature or lower  
CC in acidic buffer of pH 3.0-7.0 in the presence of a levansucrase  
CC derived from Zymomonas mobilis to produce levan, purifying the  
CC levan from the reaction solution, and subjecting it to reaction at  
CC 25-30 degree C for 3-10 hours in acidic buffer of pH 3.0-7.0 in  
CC the presence of levan fructotransferase, preferably obtained by  
CC cultivation of E. coli JUD81. The product is useful as a  
CC low-calorie sweetener.  
XX XX  
SQ Sequence 2000 BP; 337 A; 813 C; 587 G; 263 T; 0 other;  
Query Match 100.0%; Score 2000; DB 22; Length 2000;  
Best Local Similarity 100.0%; Pred. No. 2.9e-310;  
Matches 2000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 CGGTGACCCCGACTTCCCTCGAGACACCGTCCCTACCGCGCGACCGCGGCGCC 61  
Db 1 CGGTGACCCCGACTTCCCTCGAGACACCGTCCCTACCGCGCGACCGCGGCGCC 60  
Qy 62 GACTGCTCTCAGCTTAGAGCGGCGCCCTCTCTCGAGGTCTTCTGTCGGGAGCGGTGAGCGA 121  
Db 61 GACTGCTCTCAGCTTAGAGCGGCGCCCTCTCTCGAGGTCTTCTGTCGGGAGCGGTGAGCGA 120  
Qy 122 CTGCGTCGAACCTGGTCTCTCTGGGGCGCGCGGTGTGACCGCGAGCTCGAGACGCGAC 181  
Db 121 CTGCGTCGAACCTGGTCTCTCTGGGGCGCGCGGTGTGACCGCGAGCTCGAGACGCGAC 180  
Qy 182 GGCAGGAACCGTCACGTGACCGCGATGACGTCGAGGCGCCCGAGGTCTGACGCGCC 241  
Db 181 GGCAGGAACCGTCACGTGACCGCGATGACGTCGAGGCGCCCGAGGTCTGACGCGCC 240  
Qy 242 CTGAACCTGCGCGGTTCTGGGCTGACGAGCGTCCACCCCGACAGCTCTCTCTTAC 301  
Db 241 CTGAACCTGCGCGGTTCTGGGCTGACGAGCGTCCACCCCGACAGCTCTCTCTTAC 300  
Qy 302 GCTGCCGGAACAGGTTGGAGCTTCTGGGCGCCCGCTCCACGAGAGAAACAGCAA 361  
Db 301 GCTGCCGGAACAGGTTGGAGCTTCTGGGCGCCCGCTCCACGAGAGAAACAGCAA 360  
Qy 362 TGACGCGGCGCATCTCAGCGCGCGCTGTCTCCAGGAGCGCGCGCGAGACTCGGCC 421  
Db 361 TGACGCGGCGCATCTCAGCGCGCGCTGTCTCCAGGAGCGCGCGCGAGACTCGGCC 420  
Qy 422 TGATCTTTCGGCGGTGTGTGCGGCTTCAGCCCGGGGATCCGCTCCGGGTCGCTCGGTG 481  
Db 421 TGATCTTTCGGCGGTGTGTGCGGCTTCAGCCCGGGGATCCGCTCCGGGTCGCTCGGTG 480  
Qy 482 CCGTCTACACATGACGCGCGCGCTGCTGCTGAGACCGCCCGGTCACCA 541  
Db 481 CCGTCTACACATGACGCGCGCGCTGCTGCTGAGACCGCCCGGTCACCA 540  
Qy 542 CCCACGCGCTACAGCTGTACTACTCTCCGACAGAAACAGCGCGCGCGGCT 601  
Db 541 CCCACGCGCTACAGCTGTACTACTCTCCGACAGAAACAGCGCGCGCGGCT 600  
Qy 602 GGGACACGCGAGCAGCGCGCTGCTCTTACGACCGGCGCGCGCTGATGCGCG 661  
Db 601 GGGACACGCGAGCAGCGCGCTGCTCTTACGACCGGCGCGCGCTGATGCGCG 660  
Qy 662 TGGGCGCGGACTTCCCGTGTGTCGGGTGTCGGGTCGTCGGACCGCGAAGCGCAG 721  
Db 661 TGGGCGCGGACTTCCCGTGTGTCGGGTGTCGGGTCGTCGGACCGCGAAGCGCAG 720  
Qy 722 GGTTCGGCGCGCGGCTGTCGCGTCCGACCGCGCGCGCTCGGCAAGT 781  
Db 721 GGTTCGGCGCGCGGCTGTCGCGTCCGACCGCGCGCGCTCGGCAAGT 780

QY 782 ACCAGGAGTACCTCTACTGTCACGACGCGGGGTTACGTTACCGCCCTGCCCG 841  
Db 781 ACCAGGAGTACCTCTACTGTCACGACGCGGGGTTACGTTACCGCCCTGCCCG 840  
QY 842 ACCCGCTCATCGTCAACACCGAGCGTCCGCGCCACACGCGCGCGGAGATCGAGAACG 901  
Db 841 ACCCGCTCATCGTCAACACCGAGCGTCCGCGCCACACGCGCGCGGAGATCGAGAACG 900  
QY 902 CCGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCCCGCGGAGATGGGTCTGG 961  
Db 901 CCGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCCCGCGGAGATGGGTCTGG 960  
QY 962 TCATCGGACGACTCGCGTACGCGCGGTTCATACCTCGCGGACCTCGCGACTGGACAC 1021  
Db 961 TCATCGGACGACTCGCGTACGCGCGGTTCATACCTCGCGGACCTCGCGACTGGACAC 1020  
QY 1022 TTCGCGGCAACTTCGACTACCGAACCGCTCGCGGCGCATCGAGTCCCGGACCTGT 1081  
Db 1021 TTCGCGGCAACTTCGACTACCGAACCGCTCGCGGCGCATCGAGTCCCGGACCTGT 1080  
QY 1082 TCGAGATCACCGGACGAGCGGACACGCTGCTCGCGGCGCATGGAGCTGGAGCGCT 1141  
Db 1081 TCGAGATCACCGGACGAGCGGACACGCTGCTCGCGGCGCATGGAGCTGGAGCGCT 1140  
QY 1142 AGGCGATCGCGCTCCCGATGAGTACGCTACTGGACAGGACCTGGGAGCGGAGCACT 1201  
Db 1141 AGGCGATCGCGCTCCCGATGAGTACGCTACTGGACAGGACCTGGGAGCGGAGCACT 1200  
QY 1202 TCACGCGGACGACTTCACCCCGCAATGCTCGACTGGGCGTGGGACTGGTACGCGGCG 1261  
Db 1201 TCACGCGGACGACTTCACCCCGCAATGCTCGACTGGGCGTGGGACTGGTACGCGGCG 1260  
QY 1262 TCACCTGGCCATCGATCGAGCGCGCGGACCAAGCGCTCGCCCATCGCGTGGATGAACA 1321  
Db 1261 TCACCTGGCCATCGATCGAGCGCGCGGACCAAGCGCTCGCCCATCGCGTGGATGAACA 1320  
QY 1322 ACTGAGAGTACCGCGACGCGAGCTGCCACCGAGCGATCCGACGCGTACACGGGCGAGA 1381  
Db 1321 ACTGAGAGTACCGCGACGCGAGCTGCCACCGAGCGATCCGACGCGTACACGGGCGAGA 1380  
QY 1382 ACTGATCTCGCGGAGCTGCGGCTCGCCGACAGCTGGCGGCTGGTACACCTCTCTGA 1441  
Db 1381 ACTGATCTCGCGGAGCTGCGGCTCGCCGACAGCTGGCGGCTGGTACACCTCTCTGA 1440  
QY 1442 GCACCCCGCTGGAGCGCTGACGAATACTAGTCAACCGCACCAACACACTCCCGGACCGGA 1501  
Db 1441 GCACCCCGCTGGAGCGCTGACGAATACTAGTCAACCGCACCAACACACTCCCGGACCGGA 1500  
QY 1502 CCGTCGAGCGGACGCGCTTCCTGCGATGGAACGGAGCGGCGATACGAGCTCGACA 1561  
Db 1501 CCGTCGAGCGGACGCGCTTCCTGCGATGGAACGGAGCGGCGATACGAGCTCGACA 1560  
QY 1562 TCGCTGGGACACCGCGACGCTCGGATCTCGGTGGGCGCTCCCGGACGGAACCC 1621  
Db 1561 TCGCTGGGACACCGCGACGCTCGGATCTCGGTGGGCGCTCCCGGACGGAACCC 1620  
QY 1622 GGCACAGCAATCGGCAAGTACGAGGACAGCTTACGTGACCGAGGACCTCCCGACC 1681  
Db 1621 GGCACAGCAATCGGCAAGTACGAGGACAGCTTACGTGACCGAGGACCTCCCGACC 1680  
QY 1682 TCGCGGGGTACTCGTCCGCGCTTACTCGGAGCGCGCGCCCATCGACCCCGCGGCC 1741  
Db 1681 TCGCGGGGTACTCGTCCGCGCTTACTCGGAGCGCGCGCCCATCGACCCCGCGGCC 1740  
QY 1742 GATCGGTGCACTCGCGCATCTCTGTCGACACCGAGCGGTGAGGTCTTCGTCAACGCGG 1801  
Db 1741 GATCGGTGCACTCGCGCATCTCTGTCGACACCGAGCGGTGAGGTCTTCGTCAACGCGG 1800  
QY 1802 GCCACACGCTGCTCTCCAGAGGTTCACATTCGCGGAGGGGACACGGAATCTCGCTCT 1861  
Db 1801 GCCACACGCTGCTCTCCAGAGGTTCACATTCGCGGAGGGGACACGGAATCTCGCTCT 1860  
QY 1862 ACACCGGAGCGGCGCGCCGACACATTCACCGGCGATCGTCTCGCGGAGATTGGCCAGGGA 1921

Db 1861 ACACCGAGCGCGCGCCGACACTTCACCGGCGATCGTCTCGGAGATTGGCCAGGCCGA 1920  
QY 1922 TCTAGCGGATGACACACCGCGCTCACCGAAGCGCGCGCCCGGAGAGCGAGCGCGGACA 1981  
Db 1921 TCTAGCGGATGACACACCGCGCTCACCGAAGCGCGCGCCCGGAGAGCGAGCGCGGACA 1980  
QY 1982 ATCGACACGCTCTCGTCTGTT 2001  
Db 1981 ATCGACACGCTCTCGTCTGTT 2000  
RESULT 3  
ABL54835  
ID ABL54835 standard; DNA; 1752 BP.  
XX AC ABL54835;  
XX 01-JUL-2002 (first entry)  
XX Levan fructotransferase related DNA sequence.  
XX DE Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;  
XX KW levan; gene; ds.  
XX OS Arthrobacter nicotinovorans.  
XX PH Key Location/Qualifiers  
XX CDS 4..1461  
XX FT /\*tag= a  
XX FT /product= "levan fructotransferase related protein"  
XX JP2002017366-A.  
PN 22-JAN-2002.  
PD 06-JUL-2000; 2000JP-0205756.  
PF 06-JUL-2000; 2000JP-0205756.  
PR (NIPT ) NIPPON TENSU SEITO KK.  
PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.  
XX WPI; 2002-287313/33.  
XX A new levan fructotransferase -  
XX Example 1; Fig 4; 15pp; Japanese.  
XX The invention relates to a new levan fructotransferase (LFTase). LFTase  
CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to  
CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention  
CC is used for the preparation of DFA IV. The current sequence represents  
CC a levan fructotransferase related DNA sequence.  
SQ Sequence 1752 BP; 351 A; 592 C; 506 G; 303 T; 0 other;  
Query Match 47.2%; Score 945.2; DB 24; Length 1752;  
Best Local Similarity 74.3%; Pred. No. 2.9e-142;  
Matches 1237; Conservative 0; Mismatches 413; Indels 14; Gaps 3;  
QY 314 AGGTGGAGCTTCCTCGCGCCACCGCTCCACGAGAGAAC--AGCAATGACGCCGCGC 371  
Db 42 AGCTTGGCGTGTCTCGCGCGCGCGCAACTGAGAGGAACGAATCGATGACGTATGA 101  
QY 372 CATCTCAGCGCGCGCTGCTCCAGGAGCGCGCGCGGACACTCGCCCTGATCTCGG 431  
Db 102 CATCTCAGCGCGCGCTGCTCCAGGAGCGCGCGCGGAGGTTGGTCTGGACATTTTCATGAG 161  
QY 432 CGGTGCTGCGCGCTGCGAGCGCGCGGCGATCGCTCCGGGCTCGCTCCGCGCTACCA 491  
Db 162 CAATGCCATTCCCGTGGCGCGCCA-----CGCCGAGGCGATCCCTCCGGGCGATCTACCA 215

[illegible]

|                |      |  |      |
|----------------|------|--|------|
| D <sub>b</sub> | 1290 | CACGGCAGCAACGTGGAGTCTCGGTTGGCCGCCTCGTCGATGGCAGCGGCATACGAA    | 1349 |
| O <sub>y</sub> | 1632 | CATCGCAAGTACGGAGCAGACTGTACGTCGACCAGGACGCTCCGACCTCGCCGGGTA    | 1691 |
| D <sub>b</sub> | 1350 | CATCGCAANATCGGTGACGAGTTGACGTGATCGCATCTCCGGAGCAAAGCGTTA       | 1409 |
| O <sub>y</sub> | 1692 | CTCGTCGCCGCCCTACTCGCGGACGCGCGGCCCATCGACCCGCGGGGCCGATCCGTGCA  | 1751 |
| D <sub>b</sub> | 1410 | TGCGCTGGCACCTTACACCGCGCCGCCGCCCATCGATGCGAACCGACATCCGTCCA     | 1469 |
| O <sub>y</sub> | 1752 | CCTGCGCATCTCGTGACACACCCAGAGCGTCGAGGTCTTCGTCACGCGCGGCCACACCGT | 1811 |
| D <sub>b</sub> | 1470 | CCTGCGCATCTTTGTAGACACCAAGTGTGAGGTGTTCTGTAATTCCGGGCACACGTT    | 1529 |
| O <sub>y</sub> | 1812 | GCTCTCCAGCAGTCTACTTCGCCGAGGGGACACGGGAATCTCGCTCTACACGCGCGG    | 1871 |
| D <sub>b</sub> | 1530 | GGTTTCGACGAGTGACATTTCGCGCCCGGGACACGGGGATCTCCCTATATGCGGACGG   | 1589 |
| O <sub>y</sub> | 1872 | CGGCCCGCACACTTACCGGCATCTGCTCGCGAGATTGGCCAGGGCGATCTAGGCGAT    | 1931 |
| D <sub>b</sub> | 1590 | CGGTCCGCCCACTTACCGGGATCACCATCGCGAGTTCGGGAACCCCATCTAGCCTG     | 1649 |
| O <sub>y</sub> | 1932 | GCACACACACCGCTCACCGAACCCGCCGCCCGGGAGACGACGG                  | 1975 |
| D <sub>b</sub> | 1650 | CGTCCCACGCGTGAAGGACGACGCGACGCTGTGACGAGGCGG                   | 1693 |

## RESULT 4

|           |  |
|-----------|--|
| AAAX19827 | AAAX19827 standard; DNA; 1551 BP.                                      |
| ID        | XX   |
| AC        | XX   |
| AC        | XX   |
| AC        | XX   |
| DT        | 10-JUN-1999 (first entry)  |
| XX        | XX   |
| DE        | Arthrobacter nicotinovorans levan fructotransferase encoding DNA #2    |
| XX        | XX   |
| KW        | Arthrobacter nicotinovorans; levan fructotransferase; ds.              |
| XX        | XX   |
| OS        | Arthrobacter nicotinovorans.   |
| XX        | XX   |
| PN        | JF11069978-A.  |
| XX        | XX   |
| PD        | 16-MAR-1999.   |
| XX        | XX   |
| PF        | 28-AUG-1997; 97JP-0232421.   |
| XX        | XX   |
| PR        | 28-AUG-1997; 97JP-0232421.   |
| XX        | XX   |
| PA        | (NIOC ) NIPPON OIL CO LTD.   |
| XX        | XX   |
| DR        | WPI; 1999-247463/21.   |
| DR        | P-PSDB; AAY04105.  |
| XX        | XX   |
| PT        | Levan fructotransferase gene - for recombinant production of levan     |
| PT        | fructotransferase  |
| XX        | XX   |
| PS        | Claim 4; Page 10-11; 14pp; Japanese.                                   |
| XX        | XX   |
| CC        | The present sequence encodes Arthrobacter nicotinovorans levan         |
| CC        | fructotransferase. The present invention also describes a method       |
| CC        | for the preparation of levan fructotransferase in which a transformant |
| CC        | is cultured in a medium and levan fructotransferase is collected from  |
| CC        | the culture. The method can prepare levan fructotransferase in a       |
| CC        | large amount.  |
| XX        | XX   |
| SQ        | Sequence 1551 BP; 315 A; 524 C; 443 G; 269 T; 0 other;                 |

Query Match 46.6%; Score 931.8; DB 20; Length 1551;  
Best Local Similarity 75.8%; Pred. No. 4e-140;  
Matches 1184; Conservative 0; Mismatches 367; Indels 12; Gaps 2

2:



XX The invention relates to a new levan fructotransferase (LFTase). LFTase  
CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to  
CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention  
CC is used for the preparation of DFA IV. The current sequence represents  
CC the levan fructotransferase encoding sequence.  
XX  
SQ Sequence 1467 BP; 304 A; 494 C; 419 G; 250 T; 0 other;  
  
Query Match 45.6%; Score 912; DB 24; Length 1467;  
Best Local Similarity 77.0%; Pred. No. 5, Re-137;  
Matches 1126; Conservative 0; Mismatches 330; Indels 6; Gaps 1;  
  
Qy 466 CCGGGCTGCTCGTGGCGTCTACACATGACGGCCCCCGGCTGCTGCGACCCC 525  
Db 7 CAGGCATCCCTCGGGCGATCTACACATGACCGCCCTCGGGCTGGCTATGTATCCG 66  
  
Qy 526 CAACGCCGCTCACCACCCAGCGGCTTACCAGCTGTACTACTTCCAGCAGAAC 585  
Db 67 CAGGCATCCCTGATACATAAAGCGGCTTACCAGCTGTACTACTTCCAGCAGAAC 126  
  
Qy 586 AACGGCCCGGCTGGGACACGCGACGACGACGCGGCTTCCAGCAGAAC 645  
Db 127 AACGGACCGGCGATGGGACACGCGACGACGCGGCTTCCAGCAGAAC 186  
  
Qy 646 GGCACGCTGATCGGCTGCGGCTGCGGCTTCCCGCTGCTGGTGGCGGCTGCTGGC 705  
Db 187 GGAGTGGTGTGATGCAATGCAACCGGCTTCCCGCTGCTGGTGGCGGCTGCTGGC 246  
  
Qy 706 ACCGGACACGCGAGGCTTCCGGCGGCGGCTGCTGGGCTGCGGACCGGAC 765  
Db 247 ACCGCCAACCGCGGCTTCCGGCGGCGGCTGCTGGGCTGCGGACCGGAC 306  
  
Qy 766 GACGGCTGCCCAAGTACCGAGGACGATCTTACTTGGTGGCGGCTTACACCTCGCGGAAC 825  
Db 307 GACG-----GAAATTCAGAGACAGTACCTTTACTTGGTCCAGGATGGCGGTACTCC 360  
  
Qy 826 TTCACCGGCTTCCCGACCGGCTGCTGCAACACCGGCTGCGGCGGCTGCGGACCGG 885  
Db 361 TTCACCGGCTTCCCGACCGGCTGCTGCAACACCGGCTGCGGCGGCTGCGGACCGG 420  
  
Qy 886 GCCAGATGAGACCGGCTGCTGCGGACCGGCTTCCCGACCGGCTTCCCGACCGGCTGCGG 945  
Db 421 GCCAGATGAGACCGGCTTCCCGACCGGCTTCCCGACCGGCTTCCCGACCGGCTGCGG 480  
  
Qy 946 GGAGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1005  
Db 481 AACAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
  
Qy 1006 CTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065  
Db 541 CTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
  
Qy 1066 GAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1125  
Db 601 GAATGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
  
Qy 1126 GCCAGATGAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185  
Db 661 GCCAGATGAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
  
Qy 1186 TGGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1245  
Db 721 TGGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
  
Qy 1246 GACTGGTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1305  
Db 781 GACTGGTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
  
Qy 1306 ATCCGCTGGATGAACACTGAGTACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365  
Db 841 ACAGCGTGGATGAACACTGAGTACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

Qy 1366 GGCTACACGGGAGAGAACTCGATCGTCGGAGCTGGGGTCTGCGCCGACACGCTTGGGGC 1425  
Db 901 GGCTATAACGGGAGAAATTCATACACGGCGAGCTCAGGCTCGAGCGGCAATTCGGGGC 960  
  
Qy 1426 TGGTACACCTTCTGAGCACCGGCTGCGAGCTGAGGAACTAGTACCGCCGACGAC 1485  
Db 961 TGGTACACCTTCTGAGCACCGGCTTCCGGCGCTTTCGAACTATGCCACCTCCAGCAC 1020  
  
Qy 1486 ACATCTCCCGGACCGGACCGCTGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545  
Db 1021 ACCCTTCGGGACCGGACGCTCAACGGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
  
Qy 1546 GAGATCGAGCTCGATCGCTGCGACACCGGCTGAGGAACTAGTACCGCCGACGAC 1605  
Db 1081 GAACTGGAACCTCGATATTCATGCGGACGACGCGGCTGAGGCTGCTGCTGCTGCTGCTGCTG 1140  
  
Qy 1606 TCCCGGACGAGAACCGGACGACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665  
Db 1141 TCGTCCGATGCGGCGGCTTACGAACTACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
  
Qy 1666 CGAGGACCTTCCGACCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1725  
Db 1201 CGCGCATCTCGAGCAAGCGGTTATGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1260  
  
Qy 1726 ATCGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1785  
Db 1261 ATCGATGCGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
  
Qy 1786 GTCCTCGTCAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1845  
Db 1321 GTGTTCTGAAATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
  
Qy 1846 ACGGGAACTCTGCTTACACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1905  
Db 1381 ACGGGATCTCTCTTATGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1960  
  
Qy 1906 GAGATGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1927  
Db 1441 GAGTTCGGGAAACCGCATCTAAG 1462  
  
RESULT 6  
AA19826  
ID AA19826 standard; DNA; 1452 BP.  
XX AC AA19826;  
XX  
XX 10-JUN-1999 (first entry)  
XX  
XX Arthrobacter nicotinovorans levan fructotransferase encoding DNA #1.  
XX Arthrobacter nicotinovorans; levan fructotransferase; ds.  
XX Arthrobacter nicotinovorans.  
XX JPI1069978-A.  
XX  
XX 16-MAR-1999.  
XX  
XX 28-AUG-1997; 97JP-0232421.  
XX  
XX 28-AUG-1997; 97JP-0232421.  
XX (NIOC) NIPPON OIL CO LTD.  
XX  
XX WPI; 1999-247463/21.  
XX P-PSDB; AAY04104.  
XX  
XX Levan fructotransferase gene - for recombinant production of levan  
XX fructotransferase  
XX  
XX Claim 2; Page 9-10; 14pp; Japanese.  
XX



CC The present sequence encodes Arthrobacter nicotinovorans levan  
CC fructotransferase. The present invention also describes a method  
CC for the preparation of levan fructotransferase in which a transformant  
CC is cultured in a medium and levan fructotransferase is collected from  
CC the culture. The method can prepare levan fructotransferase in a  
CC large amount.

XX  
SQ Sequence 1452 BP; 299 A; 491 C; 416 G; 246 T; 0 other;

Query Match 45.5%; Score 909.6; DB 20; Length 1452;  
Best Local Similarity 77.0%; Pred. No. 1.4e-136;  
Matches 1123; Conservative 0; Mismatches 329; Indels 6; Gaps 1;

QY 466 CCGGGCTCGCTCGCGCTCTACACATGACGCCGCCCGCGGCTGTGCGACGCC 525  
DB 1 CAGGCATCCCTCCGGCGATCTACACATGACGCCGCCCGCGGCTGTGCGATCG 60  
QY 526 CAAACCCCGGTACACCCACCGCGCTTACAGCTGTACTACCTGCATCTCGACAGAAC 585  
DB 61 CAGCCACCGGTACATACAAACCGCGCTTACAGCTGTACTACCTGCATCTCGCGCAGAAC 120  
QY 586 AACGCCCGCGGCTGGGACACCGAGCAGACCGGCGCTTACACGACACAC 645  
DB 121 AACGACCGCGGATGGGACACCGACCGCGGCGAGTGTCTTACACCCACCAT 180  
QY 646 GGCACCGGTATCGCGCTCGCGGCGGCTTCCCGGTGTGTGTCGGGTCTGTCGGC 705  
DB 181 GGAGTGGTATGCCAATGCAACCGGCTTCCCGGTGTGTGTCGGGTCTGCGAGTAGTGAC 240  
QY 706 ACCGGAACCGGAGGTTCGGCGCGCGCGGCTGTCGGCTCGGACCGCGGAC 765  
DB 241 ACCGGAACCGGAGGTTCGGCGCGCGCGGCTGTCGGCTCGGACCGCGGAC 300  
QY 766 GAGCGCTCCGAGTACGAGGAGTACCTTACTGTGTGTCGACCGCGGCTTACG 825  
DB 301 GAGC-----GAAATTCAGGAACAGTACCTTACTGTGTGTCGACCGGATGGCGGTACTCC 354  
QY 826 TTCACCGCGCTCGCGACCGCGTGTATGTCACACCGAGTGTGCGCGCGCGACCGCC 885  
DB 355 TTCACCGCATTCGCTGACCGCGGTATTTGTAACACTGATGGACGACGCGCACCCGCC 414  
QY 886 CCGGAGATGAGAACCGCGAGTGTTCGCGACCGCCAGATCCACTGGGACCGCGCGC 945  
DB 415 CCGGAGTGGAGAGCGAGATGTTGTCGCGACCGCGGAAATTCAGTGGAGCGGACGCGC 474  
QY 946 GAGAAATGGTCTGCGTATCGGAGGAGTGTGCGGCTTCTACACCTCGCGGAC 1005  
DB 475 AACGAGTGGTCTGTGTCATCGGAGGCGCGCTACGCTTCTACACCTTCCCGAAC 534  
QY 1006 CTGCGGAGTGGACACTTTCGCGGCACTTCGACTACCGAACCGCGCTCGGCGGCATC 1065  
DB 535 CTGCGGAGTGGCAATGGAAGTCCAACTTCGACTACCGCAACCGCGCTCGGCGGTATC 594  
QY 1066 GAGTCCCGGACCTGCTGAGATACCGACGAGGAGGACCGACCTGGTGTGTCGCC 1125  
DB 595 GAATCCCGGATCTTCTGAAATGACCGAGAGCGGACCGGACCTGGTGTGTCGG 654  
QY 1126 GCCAGCATGGAGCGCTACGCGATCGCGCTTCCCATGAGTACGCTTCTGACGAGGAC 1185  
DB 655 GCGAGCATGGAGCGCTACAGCATCGCTTGGCCATGACCTTGTGCTACTGACAGGTTCA 714  
QY 1186 TGGAGCGGAGCAGTTCACCGCGACGACCTTACCGCGCAATGCTCGACTGGGGGTGG 1245  
DB 715 TGGACGCGCAGCATTCATCGCGCAGCAACCTTACACCGACAGTGGCTTGAAGTGGGATGG 774  
QY 1246 GACTGTACGCGCGGTACCTGGCCATCGATGAGCGCGCGGACCAACCGGCTCGCC 1305  
DB 775 GACTGTACGCGCGGTACCTGGCCCGGTGGGAGCACCTTGAGCAACCGGCTTGGC 834  
QY 1306 ATCGCGTGGATGAACAACTGGAAGTACCGCGCAGCGAGTCCCGCACCGAGCATCCGAC 1365  
DB 835 ACAGCGTGGATGAACAACTGGAATATCGCGCGCGCAACGTTGCCCGACGCGGCTCGAT 894

QY 1366 GGCTACAACGGGCGAGAACTCGATCTCGCGAGCTGGCGCTCGCCCGACAGCGCTGGCGGC 1425  
DB 895 GGCTATAACGGGCGGCAAAATTCATACGCGCGAGCTCAGGCTCGAGCGCAATCGGCGGC 954  
QY 1426 TGGTACACCTCTCAGCACCGCGCTGGAGCGCTGAGCACTAGCTACCGCCACCGAC 1485  
DB 955 TGGTACACCTTGTCTCAGCACCGCGCTTCCGGCGCTTCCGAACTAGCCACCTCCAGCAC 1014  
QY 1486 ACATCCCGCGGACCGGCTCGACGCGAGCGCTCTCGCCATGGAACGAGCGCATAC 1545  
DB 1015 ACCCTTCGGACCGCAGCTACAGCGAGTTTCGTACTTCCGTGGAGCGGCTGGGCGTAT 1074  
QY 1546 GAGATCGAGCTCGACATCCCTGGGACACCGGACAGCAAGTCCGGTGGGCGGC 1605  
DB 1075 GAACCTGGAACCTGATATTTTCATGGGACACGCGCAGCAAGCTCTCGGTGGGCGGC 1134  
QY 1606 TCCCGCGGAGGAAACCGCGGACACGACATCGCAAGTACGGGAGCAGACCTGTACGTCGAC 1665  
DB 1135 TCCTCGGATGGCGCGCATACGAACTCGGCAATACGGTGAAGAGTTGTACGTCGAT 1194  
QY 1666 CGAGGACCTTCGACCTCGCGGCTACTCTCGCGCGCTTACTCGGAGCGCGCGCGCC 1725  
DB 1195 CGCGATCTTCGAGCAAGCGGTTATGGCTGGCACCTACACCGCGCGCGCGCGC 1254  
QY 1726 ATCGACCGCGCGCGCGATTCGCTGACCTGGCATCTCTCGTGGACACCGCAGAGCTGAG 1785  
DB 1255 ATCGATGGAACGCGCAGATCGCTCCACCTGGCATCTTTGTAGACACCAAGTTGTGAG 1314  
QY 1786 GCTTCGCTCAACGCGCGGACACCGCTCTCCAGAGGTCACCTTCGCGGAGGCGAC 1845  
DB 1315 GTGTTCTGAAATTCGCGGACACGCTGTTTCGAGAGGTCGCTTCGCGCGCGGCGAC 1374  
QY 1846 ACGGAAATCTCGCTACACCGACCGCGCGCGCGCACACTTCACCGGATCTCTCGCGC 1905  
DB 1375 ACGGGATCTCCCTCTATCGGAGCGCGCTCGCGCAACTTCACCGGATCATCATCCG 1434  
QY 1906 GAGATTGCCAGCGCATC 1923  
DB 1435 GAGTTCGGGAACCCCATC 1452

RESULT 7  
AAD17184  
ID AAD17184 standard; DNA; 65140 BP.  
XX  
AC AAD17184;  
XX  
XX  
DT 29-NOV-2001 (first entry)  
XX  
DE Streptomyces noursei nysl DNA of nystatin PKS gene cluster.  
XX  
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; nysl; ds.  
XX  
OS Streptomyces noursei.  
XX  
FH Location/Qualifiers  
FT CDS complement (1..1035)  
FT /tag= a  
FT /product= "NysD2 partial protein"  
FT /note= "CDS does not include stop codon"  
FT CDS complement (1056..2576)  
FT /tag= b  
FT /product= "NysD1 protein"  
FT 2806..6906  
FT /tag= c  
FT /product= "NysA protein"  
FT 6952..16530  
FT /tag= d  
FT /product= "NysB protein"  
FT 16550..49840  
FT /tag= e  
FT /product= "NysC protein"



```
FT CDS 50260..51015
FT /tag= f
FT /product= "NysE protein"
FT 51405..54305
FT /tag= g
FT /product= "NysR1 protein"
FT 54329..57190
FT /tag= h
FT /product= "NysR2 protein"
FT /note= "CDS does not include start codon"
FT 57180..59963
FT /tag= i
FT /product= "NysR3 protein"
FT 60415..61047
FT /tag= j
FT /product= "NysR4 (short) protein"
FT /note= "CDS does not include start codon"
FT 61736..62497
FT /tag= k
FT /product= "NysR5 protein"
FT /note= "CDS does not include start codon"
FT complement (62551..63615)
FT /tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT 63765..64961
FT /tag= m
FT /product= "ORF1 protein"
FT
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYN0-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
XX P-PSDB: AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
XX
XX Query Match 7.4%; Score 147.6; DB 22; Length 65140;
XX Best Local Similarity 45.4%; Pred. No. 2.3e-15;
XX Matches 867; Conservative 0; Mismatches 1004; Indels 37; Gaps 8;
```

```
QY 8 ACCCGGACTTCCTCGACGACACACCTCCCTACCGCGCGACCGCGCGCGCGCGCTGC 67
DB 19803 ACGGCGGACACCGACCGCGCGCTTCGCGACTGCACCGCGCGCTTCGCGCGCAC 19862
QY 68 TCCTCAGCCTAGAGGGGCGCTCTCGAGGTCTTCGTGGGGAGGCTGAGGCGACTCGCT 127
DB 19863 AACACGCGCGCGCTACGCGGAGCTCGCGGCCATACGCGCGCGCGCTTCGCGCT 19922
QY 128 CGAACCTGGTCTCTCGGGGCGCGGCTGTGACCGGAGCTTCGAGAGCGGCGGCGAG 187
DB 19923 GGAAGGCGTCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19982
QY 188 GAACCGTGCACGTGACCGCGATCGACGTGAGGCGCGCGCGCGCGCGCGCGCGCG 247
DB 19983 CCGCGGAGGACACCGTCAACCATCGCGTCTACGACGCGCGCGCGCGCGCGCG 20042
QY 248 CTGCGCGCGTTCGGGCTGAGGAGCGCTCCACCGCGCGCGCGCGCGCGCGCGCGCG 307
DB 20043 TCGACTCCCTGGTCTCCCGGAGGTCCCGCGGAGTCCCGCGGCGCGCGCGCGCG 20102
QY 308 CGAACGAGGTGGACGCTTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 367
DB 20103 ACGCGGACTCCCTCTTCACGTGAGTGGACCGCGCTCCAGGGCGCGCGCGCGCG 20162
QY 368 CGGCAATCTCACGCGCGCGTGTCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCT 427
DB 20163 CGGCCA-----CGTGCCTGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 20213
QY 428 TCGGCGGTGTGTGCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 487
DB 20214 TCGCG-----CCACCGGATCCGGACACCGCGCGCGCGCGCGCGCGCGCGCT 20264
QY 488 ACCACATGACGCGCGCGCGCTGTCTGCGACCGCGCGCGCGCGCGCGCGCGCGCG 547
DB 20265 CCGACGCGGAAAGCGCGCTCCCGACCTGTGTGTCACACCTCTAC-----CAC 20321
QY 548 GCGCCTACCAAGCTGTACTACTCTGCACTCCGACACAGAACACGCGCGCGCGCG 607
DB 20322 GCGCGCGCGTCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20381
QY 608 ACGGAGACGACGCGCGCGGTGCGCTTACGACACACGCGCGCGCGCGCGCGCGCG 667
DB 20382 AGTGGCTCGCGGACGCGCTTCGCGGACCGCGCGCGCGCGCGCGCGCGCGCG 20441
QY 668 CCGACTTCCCGGTGTGTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 727
DB 20442 CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20501
QY 728 GCGCGCGCGCGGTGTGTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
DB 20502 AGAACCGCGCGCGTTCGCGCTCTCTCGACCTCGCGCGCGCGCGCGCGCGCG 20561
QY 788 AGCAGTACCTCTACTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 847
DB 20562 AGACCTTGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20618
QY 848 TCATCGTCACACGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 907
DB 20619 GCACCGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20678
QY 908 GGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 967
DB 20679 GGAACCGGACGCGCGCGCTCTGATCACGCGCGCGCGCGCGCGCGCGCGCG 20738
QY 968 GAGCAGTGGGTAGCGCGCGCTTCTACACCTCGCGCGCGCGCGCGCGCGCGCG 1027
DB 20739 CCGCGACCTTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20798
QY 1028 GCAACTTCGACTACCGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
DB 20799 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20858
```







Db 103863 CGGCCGAGCGCCGCAVAGACGGNNHNNHNSGCCCGCCGCGCAVAGCGCNCNNHNN 103804

Qy 1454 CAGCGCTGACGAATAGCTACCGCCACACACACTCCCGACCGGACCGCTCGACGGCA 1513

Db 103803 NSGCCGCGCGGACCGCCGCAVAGACGGCNCNNHNNHNSGCCCGCCGCGCAVAGCG 103744

Qy 1514 GCGCGCTCTGCTATGTAAGACGCGCGCATACGATGAGATCGAGCTCGATCGCTGGGACA 1573

Db 103743 GCCNNHNNHNSGCCCGCCGCGCAVAGACGGCNCNNHNNHNSGCCCGCCGCGCAVAG 103684

Qy 1574 CCGCGAGCAAGCTGGCATCTCGTGGCGCGCTCCCGGACCGGAACCGCGCACAGACA 1633

Db 103683 CCGCAVAGACGGCGCNCNNHNNHNSACGGCGCNCNNHNNHNSACGGCGCNCNNHNSV 103624

Qy 1634 TCGCAAGTAGGAGACAGCTGTACCTCGACCGAGGACCTCCGACCTCGCCGGGTACT 1693

Db 103623 GACGGCGCNCNNHNNHNSAVGACGGCGCNCNNHNNHNSCAVAGCGCGCGCNCNNHNSC 103564

Qy 1694 CGCTGCGCCCTACTCGGAGCGCGCGCCGCCCATCGACCGCGCGCGCATCGCTGCAAC 1753

Db 103563 CAVGACGGCGCGCNCNNHNNHNSGCGCAVAGCGCGCNCNNHNNHNSGCGCAVAGCGCGCG 103504

Qy 1754 TGGCATCTCTGTCGACACCGACGCTCGAGGTCTTCGTCAACGCGCGGCACACCGTGC 1813

Db 103503 CNNHNNHNSCGCCCAVAGACGGCGCGCNCNNHNNHNSGCGCAVAGCGCGCGCNCNNHNSC 103444

Qy 1814 TCTCCAGAGGTCCACTTCGCGGAGGCGGACACGGGAATCTCGCTCTACACCGACGGCG 1873

Db 103443 GCCCAVAGCGCGCGCNCNNHNNHNSACGGCGCAVAGCGCGCGCNCNNHNSGACGCC 103384

Qy 1874 GCCCGCACACTTCACCGGCATCTGCTCGCGGAGATGGCCGCGGATCTAGCGGATGC 1933

Db 103383 GCCAVGACGGCGCGCNCNNHNNHNSGCGCGCAVAGCGCGCGCNCNNHNNHNSCGGACGC 103324

Qy 1934 ACACACACCGCTCACCGAGCGCGCGCGCG 1964

Db 103323 CGCCAVGACGGCGCGCNCNNHNNHNSCGGACG 103293

RESULT 10

AAF25795

ID AAF25795 standard; DNA; 3849 BP.

XX

AC AAF25795;

DT

09-APR-2001 (first entry)

XX

DE S. chrysomallus acm-C DNA.

XX

KW Peptide synthetase; PPS; actinomycin synthetase; acm-C; penicillin;

KW N-methyltransferase domain; activation domain; amino acid synthesis;

KW vancomycin; cephalosporin; pristamycin; actinomycin D; ds.

XX

OS Streptomyces chrysomallus.

XX

PN DE19928313-Al.

XX

PD 21-DEC-2000.

XX

PF 16-JUN-1999; 99DE-1028313.

XX

PR 16-JUN-1999; 99DE-1028313.

XX

PA (KELL/) KELLER U.

XX

PI Schauwecker F, Keller U;

XX

DR WPI; 2001-081744/10.

XX

PT Introducing N-methyltransferase activity into peptide synthetase

PT enzymes, useful for synthesis of N-methylated peptides such as

PT antibiotics, by altering DNA

XX

## Example 2; Page 9; 18pp; German.

PS This invention describes a novel method of introducing an  
XX N-methyltransferase domain (A) into peptide synthetase (PPS) activation  
CC domains by altering the DNA that encodes the domains. The invention also  
CC describes a method for combining genes (or gene segments) that encode PPS  
CC modules that lack (A) with genes (or segments) encoding modules that  
CC include (A). Modified DNA sequences formed by insertion of (A) are used  
CC (i) for altering natural (or already altered) PPS or polyketide  
CC synthetase (PKS) genes and their fragments, also for construction of new  
CC PPS and PKS gene and (ii) for construction of plasmids or genetically  
CC altered organisms for synthesis of encoded proteins (B). (B) are used for  
CC in vivo or in vitro enzymatic synthesis of amino acids, polypeptides, and  
CC peptidyl-acetyl mixed structures containing N-methylated amino acids, or  
CC their derivatives, also for fermentative production of such compounds.  
CC These compounds are often of pharmaceutical value, e.g. penicillin,  
CC vancomycin, cephalosporin, pristamycin or actinomycin D. The modified  
CC PPS (enzymes) involved in non-ribosomal peptide synthesis) are able to  
XX N-methylate their substrates but substrate specificity remains unchanged.

SQ Sequence 3849 BP; 596 A; 1685 C; 1122 G; 446 T; 0 other;

Query Match 6.0%; Score 119.8; DB 22; Length 3849;  
Best Local Similarity 44.6%; Pred. No. 7.9e-11;  
Matches 558; Conservative 0; Mismatches 687; Indels 6; Gaps 2;

Qy 583 AACAAACGCGCCGCGGTGGGACACGACGACGACGACGCGCTTCACGCAC 642

Db 192 AGCAGCGCGCCAGCTCAGCTACTCGAGCTCAACAGCGCGCAACCACTCGCCAC 251

Qy 643 CACGGCACCGTGTATCGCGCTGCGCGCGCTTCCCGCTGTGTGTGCGGGTGGCGGTGCT 702

Db 252 CAGCTACCAACCGCGCGCATCGCGCGCGGACGCGCTCGCGCTCTCTCAACGCTCC 311

Qy 703 GGCACCGGACACGCGGAGGTTCGCGCGCGCGGTGCTGCGCTGCGGACCGCACCG 762

Db 312 CCCGACACGCTCACCACTCGCTCTCGCCCAAGACGCGCGGACCTACATCCCCCTC 371

Qy 763 ACCGACGCGCTCCGCAAGTACGAGGAGCTACTCTACTGTGTCGACGCGCGGTTC 822

Db 372 GACAGCGCTACCGCGCGCGCTACCGCTCGCTCGACGAGACCGCGCAAACTC 431

Qy 823 AGTTTACCGCGC---TGCGCGACCGCTGCTATCTCAACACGCGGTGCGCGCGCGCAC 879

Db 432 CTCATCACCGACACACGACCTCGACACCAACCAACAGTTCAACCGCGCGGAC 491

Qy 880 ACGCGCGCGGAGATCGAGACGCGGAGTGTTCGCGACCGCAAGATCCACTGGGACAC 939

Db 492 ACCCGCGCGGAGAGACCGCGCGCAACCGCAACCAACCGCGCGCGGACCGCGC 551

Qy 940 GCGCGCGGAGAAATGGTCTGCTGCTCATCGGACGCTGCGGTACGCGCGGTTCACACCTCG 999

Db 552 GCCTATCATGTACACGAGGCTCCACCGCGCGCGCGCGCGGTATCGCGCACCCAC 611

Qy 1000 CCGAACTGCGCGACTGGACACTTCGCGCGCACTTCGACTACCGACCGCGCTCGCG 1059

Db 612 CGCAACATCACCGCGCTCGCGCTCGACCGCGCTTCGACCGCGCGCGCGCGGTG 671

Qy 1060 GGCATCGAGTCCCGGACCTTTCGAGATCACCGGACGCGGACCGCGCGCGGTG 1119

Db 672 CTCCTCCATCCCGCGCGCGCTTCGACGCGCTCTACCTACGAGATCTGGTCCCGCTCTC 731

Qy 1120 CTCGCGCGCGGACGCGCTACGCGCTCGCGCTCCCGCTACGCTACGCTACTGGACA 1179

Db 732 AACGGCAACCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791

Qy 1180 GGCACCTGGGACGCGGAGGATTCACCGCGGACGACTCACCGCGCAATGGCTGACTGG 1239

Db 792 GTCATCACCGACGAGATCACCGCGCTCTGGCTGACGAGCTGGTCTTCAACTCTCTC 851

Qy 1240 GGTGCGGATGTTAGCGCGCGGTTCACCTGCGCATCGATCGACGCGCGCGCGCGCG 1299

Db 852 ACCGAGCAGAGCGCGGAGACCTTCACCGCGGTTCGCGGAGATCTGGACCGCGCGGAGCC 911







Db 21093 CTGCGCGCGCGCGCGCGCTACCGAGAGTCTCTCGCCACCATCCGCGCGCGCG 21152  
QY 1689 GTACTCGTGGCGT 1748  
Db 21153 CGACGAACCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21212  
QY 1749 GCACCTGGCG 1808  
Db 21213 GCGCGTGTCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21272  
QY 1809 CGTGTCTTCCG 1868  
Db 21273 CGCGTACACATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21332  
QY 1869 CG 1880  
Db 21333 CG 21344

## RESULT 12

AAD17185

ID AAD17185 standard; DNA; 27541 BP.

XX

AC AAD17185;

XX

DT 29-NOV-2001 (first entry)

XX

DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

XX

KW antifungal; antibiotic; nys2; ds.

XX

OS Streptomyces noursei.

XX

FH Key Location/Qualifiers

FT CDS complement (454..1191)

FT

FT /\*tag- a

FT /product- "NysF protein"

FT /note- "CDS does not include start codon"

FT complement (1275..3092)

FT /\*tag- b

FT /product- "NysG protein"

FT complement (3070..4824)

FT /\*tag- c

FT /note- "CDS does not include start codon"

FT /product- "NysH protein"

FT 5122..6156

FT /\*tag- d

FT /product- "NysD3 protein"

FT 6338..27541

FT /\*tag- e

FT /product- "NysI partial protein"

FT /note- "CDS does not include stop codon"

XX

PN WO200159126-A2.

XX

PD 16-AUG-2001.

XX

PF 08-FEB-2001; 2001WO-GB00509.

XX

PR 08-FEB-2000; 2000GB-0002840.

XX

PR 10-APR-2000; 2000GB-0008786.

XX

PR 14-APR-2000; 2000GB-0009387.

XX

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

PA (SNF) SINTEF STIFTELSEN IND TEK FORSK.

PA (ALPH-) ALPHARMA AS.

PA (SINV-) SINVENT AS.

PA (DZIE/) DZIEGLEWSKA H.

PA (ZOTC/) ZOTCHEV S B.

PA (SEKU/) SEKUROVA O N.

PA (FJAE/) FJAERVIK E.

PA (BRAU/) BRAUTASET T.

(STRO/) STROM A R.

XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX

DR WPI: 2001-557614/62.

DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.

XX

PT New nystatin polyketide synthase polynucleotides and polypeptides,

PT useful as antibiotics and antifungals -

XX

PS Claim 2; Page 151-166; 266pp; English.

XX

CC The present invention relates to the cloning and sequencing of the gene

CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present

CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX

SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 5.5%; Score 109.2; DB 22; Length 27541;

Best Local Similarity 43.5%; Pred No. 3.2e-09;

Matches 859; Conservative 0; Mismatches 1088; Indels 27; Gaps 7;

QY 4 GTGCACCCCGACTTCCCTTCGAGCACACCGTCCCTACCGCGCGCGCGCGCGCG 63

Db

QY 64 CTGCTCTCAGGCTAGACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 123

Db

QY 124 GCGTCGAACCTGTCTCTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 183

Db

QY 8405 GCGTGCACG 8464

QY

QY 184 CCAGGAACCGTGACGTGACCGGATGCGGTCCAGGCGCGCGCGCGCGCGCG 243

Db

QY 8465 CTGCTCG 8524

QY

QY 244 GAACCTGCG 303

Db

QY 8525 GTCTGTCTCG 8584

QY

QY 304 TCCCGCAACAGGCTGACGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 363

Db

QY 8585 CGCAAGACCG 8644

QY

QY 364 ACGCCGGCGATCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423

Db

QY 8645 CTGACGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8704

QY

QY 424 ATCTTCGGCGGTGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 483

Db

QY 8705 GTCTTCAACCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8764

QY

QY 484 GTCTACACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 543

Db

QY 8765 GTGACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8824

QY

QY 544 CAGGGGCGCTACAGTGTGA---CTACCTGCACTCGGACCAAGAACCGCGCGCG 600

Db

QY 8825 GACACCACTGCTTCTCGAATCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 8884

QY

QY 601 TGGGACCACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660

Db

QY 8885 TGCTGGACG 8944

QY

QY 661 CTGCGGCG 714

Db

QY 8945 GAGGAGCACACCTTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9004



QY 715 AGCGAGGTTGCGGCGCGGGTCTGTCGGCTCGCGAGCCACCGACCGAGCGGCGT- 773  
Db 9005 TGGACCGGCTGTTGCGCGGACCGCGCGCGCGCGACACCGACTGCCGACCTACGCTTC 9064  
QY 774 -----CCGCAAGTACACGAGAGTACCTCTACTGGTGCAGCGGCGGGTTACGTTTC 828  
Db 9065 CAGCGCGCGCGTACTGCGCCAGCGCTTCCAGAGCGCACCGCGGACCTCGGCTCGGTC 9124  
QY 829 ACCGCGCTCGCGCGCGCTCATGTC AACACCGACGCTCGCGCGCGCACCGCGCGCC 888  
Db 9125 GGCCTCGGTGCGCGCGCACCGCTGCTCTCCGCGCGCTCTCCCTCGCGCGACGCG 9184  
QY 889 GAGATCGAAGAGCGGAGTGTTCGCGAGACCGACCGAAGATCCACTGTGGACACCGCGCGCGA 948  
Db 9185 GGCACCTCTGCTACGCGCGCGCTTCCCGGAGAGACCGACCGCTCGCGCGAAC- ---- 9239  
QY 949 GAATGGGTCTCGCTCATCGGAGGACTGCGGTACGCGCGGTTCTACACCTCGCGGACCTG 1008  
Db 9240 -ACACCGTCCGCGGACCACTGCTGCGCGGTACCGCGCTTCTCGAACTCGCGGTCGCG 9298  
QY 1009 CGGACTGGACACTTCGCGCGCAACTTCGACTACCGCGAACCAGCCCTCGCGCGGATCGAG 1068  
Db 9299 CGCGCGACGAGTGGCTGCGACCGCTCGAGGAACTCACCTCGCGCGACCGCTCTCTG 9358  
QY 1069 TGCGCGGACTTTCGAGATCACCGAGACGAGCGGACACGCGCACTGGGTGCTCGCGCGC 1128  
Db 9359 CTGCGCGAAGCGCGCGCTCAGTTCAGTTGTGATCGCGCAACCGCGGAGCTGTCCGT 9418  
QY 1129 AGCATGGAGCGCTACGGCATCGGCTCCCATGATGACTAGGCTACTGGACGACCTGG 1188  
Db 9419 CGCGCACCGTCAAGCTCCAGCGCGCGCGCGCGACCGCGGAGACACCGCTTGACCGCC 9478  
QY 1189 GACGCGGAGCACTTCACCGCGGAGCACTCACCGCGCAATGGCTCGAGTGGGCTGGAC 1248  
Db 9479 CACGCGACGGGCTCTCACACCGCGGACGCTCCCGCGAGCTC-CCGCTTCGTCGCG 9536  
QY 1249 TGGTACGCGCGCTACCTGCGCATCGATGACGCGCGCGCGAGACCAAGCGCTCGCGCATC 1308  
Db 9537 AGCAGGCGCGACCGCGCTCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGTGGC 9596  
QY 1309 CGGTGGATGACAACTGGAATAGCGCGGACCGCGGACGCTCCCGCGGAGCATTCGACGCG 1368  
Db 9597 CCGCGCGCGCGCGGACCGCTGCGCTGAGCGCGGACGCTACGACGCGGCTC-GCGGACGCG 9655  
QY 1369 TACAACGCGGAGAACTCGATCGTCCGCGAGCTCGCGCTCGCGCGCGCGCGCGCGCGTGG 1428  
Db 9656 GCGTTCGCTACGCGCGCGCTTCCAGGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCTC 9715  
QY 1429 TACACCTCTGAGACCGCGCGGAGCGCTGAGCACTAGCTACGCGCGCGCGCGCGCGCGCA 1488  
Db 9716 GTCTACGCGGAGTCCGAGCTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 9775  
QY 1489 CTCGCGGAGCGGAGCGTCCGCGGAGCGCGCGCTGCGCATGGAAGCGCGCGCGCATACGAG 1548  
Db 9776 CACCGCGCGCTGCTGAGCGCGCGCGCTGCGACCGCGCGCGCTTACCGCGCGCGCGGAGCG 9835  
QY 1549 ATCAGCTCGACATCGCTGCGGAGACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGTCC 1608  
Db 9836 GCGCGGCGCGCGCTGCGCTTCTCGGAGGCGCTCTCCCTCCACGCGCGCGCGCGCGCGCG 9895  
QY 1609 CCGGAGGAAACCGCGGACGAAATCGGCAAGTACGAGAGGAGACCTGAGTCTGAGCGGAG 1668  
Db 9896 ACCCTCGCGCGCTGACCGAG 9955  
QY 1669 GGACCTTCGAGCTCGCGGCTACTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1728  
Db 9956 GGCACCGCGCGCGCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10015  
QY 1729 GACCG 1785  
Db 10016 CAGCTCGACACG 10075  
QY 1786 GTCTTCGTCACACG 1845

Db 10076 GTCCAGCGACCGCGCACCGACCGCGCGCGCGCGCGCTCTCTCGCGCGCGACCGCTTCGGC 10135  
QY 1846 ACGGAATCTCGCTCTACACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1905  
Db 10136 CTGCTACCGACCGCGGATTCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10195  
QY 1906 GAGATTGCGCGCGGATCTAGCGATGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1959  
Db 10196 CGCGCGGACGCGCGGTCGCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 10249

## RESULT 13

AAI99683/C

ID AAI99683 standard; DNA; 4403765 BP.

XX AC AAI99683;

XX DT 15-JAN-2002 (first entry)

XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX OS Mycobacterium tuberculosis.

XX PN US6294328-B1.

XX PD 25-SEP-2001.

XX PF 24-JUN-1998; 98US-0103840.

XX PR 24-JUN-1998; 98US-0103840.

XX PA (GENO-) INST GENOMIC RES.

XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;

XX DR WPI; 2001-647261/74.

XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
determining the nucleotide sequence of the strain at positions in the  
genome corresponding to positions where M. tuberculosis strains CDC  
1551 and H37Rv differ.

XX PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX CC The invention relates to evaluating strain variation within and between  
different populations of the tuberculosis bacterial pathogen,  
Mycobacterium tuberculosis or related Mycobacterium by determining the  
nucleotide sequence of the first strain at positions in the complete  
sequence of the genome that correspond to positions that differ in the  
nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
H37Rv (AAI99682). The method is useful for evaluating strain variation of  
M. tuberculosis and has valuable application in the fields of  
tuberculosis genetics, epidemiology, patient treatment and epidemic  
monitoring.XX CC Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from USPTO  
at seqdata.uspto.gov/sequence.html?docID=6294328B1.

XX SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 5.0%; Score 101; DB 22; Length: 4403765;

Best Local Similarity 42.3%; Pred. No. 4.1e-08;

Matches 833; Conservative 0; Mismatches 1115; Indels 23; Gaps 4;

QY 11 CCGACTTCCTCTGACGACCACTCCCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 70

Db 3941672 CCGCCG 3941613

QY 71 TCAGCCTAGACG 130

Db 3941612 GTGCGGCGCGTCTCCGCGGCGCCCGCGGCTCCGCGCGTCCGCGGTCGCGCGGTCCCGCGTCAACG 3941553  
QY 131 ACCTGGTCTCTTGGGGGCGCGGTGTGACCGCGAGCCTCGAGACGGCAGCGCCAGGAA 190  
Db 3941552 CCGGTGCGCGGATCCCGGTGGGGTTATCGGCGCGCGCCCGCAGCACCCCGCGCGCG 3941493  
QY 191 CCGTGCACGTGACCGGATCGAGCTGAGGCGCCCGCAGCATGCTGACGCGCCCTGAACTG 250  
Db 3941492 CCTTGTCCGCGAGTCCCGCGCGCGCGCGGAGCGTTGTGTGCGCGCGGACCGGTGTA 3941433  
QY 251 CCGCGCTTCTGGGCTGACAGCGCTCCACCCCGCAGCTCTCTTCTACCGCTGCCGA 310  
Db 3941432 CCGCGGCGCGCGCGCGCGCGCGCGCGCGTCCACCGCGAAACCGGTGCGCGCG 3941373  
QY 311 ACCAGGTGAGCGCTTGTGCGCGCGCCACCGCTCCACGAGAGAAACGCAATGAGCGCG 370  
Db 3941372 GTGGCGCGCGGTGTGCGCGGTGCGCGCGCGCGCGCGGTCCCGGTCAACCGCTTG 3941313  
QY 371 CCATCTACCGCGCGGTGCTCCAGGAGCGCGCGCGCGGAGCATCTGCGCCCTGATCTTG 430  
Db 3941312 CCGCGGCGCGCGCGCGCGCGCGCGCGGTGCTGTGTGCGGATCATGCGCGCGGTGCTCCG 3941253  
QY 431 CGGTGCTGTGCGCGCTGCGAGCGCGCGCATCCGCTCCGCGCGTCCGTCGCGTCAAC 490  
Db 3941252 GTGCGGCGCGTCCCGCGCGCGCGCGCGTCCGCGCGTCCGCGCGTCCCGCGGTCAACG 3941193  
QY 491 ACATGACGCGCGCGGTGCTGCGACCGCGCATCCGCGCGGTCAACCGCGCGCGCG 550  
Db 3941192 CCGGTGCGCGGATCCCGGTGGGGTTATCGGCGCGCGCGCGCGCGCGCGCGCG 3941133  
QY 551 CCTACAGCTGTACTACGTGCTGCGAGCAGAAACAGCGCGCGCGCGCGCGCGCG 610  
Db 3941132 CCTTGTCCGCGGTGCGCGCGCGCGCGCGCGGTGCTGCGCGCGCGCGCGCG 3941073  
QY 611 CCAGACGACGCGCGGTGCTTACGACGACGCGCGCGGTGATGCGCGTGGCGCG 670  
Db 3941072 CCGCGGCGCGCGCGTGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941013  
QY 671 ACTTCCCGGTGCGCGGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730  
Db 3941012 CCGGACGCGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940953  
QY 731 CCGGCGCGGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790  
Db 3940952 CCGCGGTGCGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940893  
QY 791 ACTACTCTACTGTCGACGCGCGGTTCACGTTACCGCGCGTCCGCGCGCGCGTCA 850  
Db 3940892 CTTGGCGCGCGCGCGCGCGCGCGCGCGGTGCTGCTGTGGTG-----CCGATCATCGCG 3940839  
QY 851 TCGTCAACACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910  
Db 3940838 CCGGTGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940779  
QY 911 TCGCGACCG 970  
Db 3940778 CCGCGGTACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940719  
QY 971 GACTGCGGTACCGCGGTGCTACACCTCGCGGAACCTGCGGCGACTGAGACTTCGCGCG 1030  
Db 3940718 CCGCGCGCGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940659  
QY 1031 ACTTGGACTTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090  
Db 3940658 GCGCGCGCGGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940599  
QY 1091 CCGGAGACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150  
Db 3940598 CCGCGTCCACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940539  
QY 1151 GCCTCCCATGATGATGCGCTTCTGACAGCGCACTGGAGCGCGCGCGCGCG 1210

Db 3940538 GCCCGCGCGGTCTCGCGGTCAACGCGCTTGGCGCGCGCGCGCGCGCGCG 3940479  
QY 1211 ACAGACTTCAACCGCGAATGGCTGAGCTGGGGTGGGACTGGTACGCGCGCGTCACTGCG 1270  
Db 3940478 GCCGTCTCGGTGGCGTGAAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCG 3940419  
QY 1271 CATCATGACCGCGCGCGGAGCAAGCGCTGCGCATCGCTGGGATGACAACTGGAAGT 1330  
Db 3940418 CTACGCGCGGTGCGCGGTCAACGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGCG 3940359  
QY 1331 AGCGCGCGCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1390  
Db 3940358 GCCCG 3940299  
QY 1391 TCCGCGAGTGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1450  
Db 3940298 CCAGCG 3940239  
QY 1451 TGGGAGCGGTGACGAACTA-----CGTACCGCGCGCGCGCGCGCGCGCG 1501  
Db 3940238 GTCCGCGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940179  
QY 1502 CCGTGCAGCGCGCGCGGTCTGCGATGGAACGCGCGCGCGCGCGCGCGCGCG 1561  
Db 3940178 CCG 3940119  
QY 1562 TCGCGTGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1621  
Db 3940118 TGTCCGCGGTGCGCGGTGAGCACCGCGCGCGCGCGCGCGCGCGCGCG 3940059  
QY 1622 GCGACAC-----GAACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1675  
Db 3940058 CCG 3939999  
QY 1676 CGCGACTCGCGCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1735  
Db 3939998 CCGACCG 3939939  
QY 1736 GCGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1795  
Db 3939938 GTGCG 3939879  
QY 1796 AGCG 1855  
Db 3939878 GCG 3939819  
QY 1856 CGCTCTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1915  
Db 3939818 CCG 3939761  
QY 1916 AGCG 1966  
Db 3939760 CGGCG 3939710

## RESULT 14

AAK53491

ID AAK53491 standard; DNA; 114955 BP.

XX

AC AAK53491;

XX

DT 05-JUL-1999 (first entry)

XX

DE Human adenosine A1 receptor antisense oligonucleotide fragment.

XX

KW Antisense oligonucleotide; multiple target; antisense treatment;

KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW acute asthma; allergy; asthma; impaired respiration;

KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW chronic obstructive pulmonary disease; leukemia; lymphoma;

KW colon cancer; breast cancer; lung cancer; pancreatic cancer;







|    |       |   |       |
|----|-------|---|-------|
| Db | 62147 | GGCGTTCGGCTCCGGCGAGGACTTCGTATATGGCCCGCGGATGACCGCCGACACGCCGAT    | 62200 |
| Qy | 1371  | CAACGGCGAGAATCG---ATCTGCTCCGCGAGCTCGGCTCGCCCGACAGAGCTGGCGGCTG   | 1427  |
| Db | 62207 | GACGGCTCCGTACCGGCCCATCTGTAGGGGCTTCGCGAGGAGCGCGCGGCTCGGCCG       | 62266 |
| Qy | 1428  | GTACACCCCTCTGTAGCACACCCCGGTGGGAGGCTGTACGAATACGTCAACGCCACCAACAC  | 1487  |
| Db | 62267 | TGCGGGCGACACGTTTCGCCACAGCGCTCGCGAGCTGCCCGACGCCGACCGCGCGCGC      | 62326 |
| Qy | 1488  | ACTCCCCGACGGACCGTGCAGCGCAGCGCGTCTTCCCATTTGGAACGAGCGCGCATACGA    | 1547  |
| Db | 62327 | GCTGACCAACCTCGCTTCGGAGCGCCAGCGCGCGTCTCGGCCACCGCAGCGCTCCGA       | 62386 |
| Qy | 1548  | GATCGAGCTCGACATCGCTTGGGACACCGGACGAGAACTCGGCATCTCGTTGGSCCGCTC    | 1607  |
| Db | 62387 | GATCGCGCCGACACGAGCTTCAAGGACTTCGGCATCGACTCGCTCAACCGGATCGAGCT     | 62446 |
| Qy | 1608  | CCCCGAGGAACCCGGGCACACGACATCGGCAAGTACGGAGCAGAGC---TGTACGTCGA     | 1664  |
| Db | 62447 | GCACAACCGGCTCGCGAGGCGGACCGGGCTCGGCTGAGTGCACACGCTGGTGTTCGACCA    | 62506 |
| Qy | 1665  | CCGAGGAGCCTCGACCTCGCGGGGTACTCGCTGCGCCCCCTACTTCGGAGCGCGCGCGCCC   | 1724  |
| Db | 62507 | CCCGACACCTCGGGTCTCTCGCGCCAAAGTCCCGACACGATCTGTTCCGGCACGCGCGTGCC  | 62566 |
| Qy | 1725  | CATCGACCCCGGCGCGGATCCGTGCACTTCGCGATCTCGTCGACACCCAGAGCGTCGA      | 1784  |
| Db | 62567 | CACGCCCGCGGACGCGACCGACCCACGAGCCACTCGCGATCTGTCGGCATGGC           | 62626 |
| Qy | 1785  | GGTCTTTCGTCGAAGCGCCGGCCACACCGTGTCTCCACGACGGTCCACTTTCGCCGAGGGCGA | 1844  |
| Db | 62627 | GTCCGACTCCCGCGGGGTGCGCTCGCCGAGGAGCCTGTGGCAGCTCGTGGCGTCCGG       | 62686 |
| Qy | 1845  | CACGG' 1849   |       |
| Db | 62687 | CACCG 62691   |       |

Search completed: May 10, 2003, 07:35:05  
Job time : 4373.5 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 04:44:16 ; Search time 2890 Seconds  
(without alignments)  
11213.553 Million cell updates/sec

Title: US-09-868-328B-3

Perfect score: 2001

Sequence: 1 ggggtgaccccgactccc.....atcgacacgtctctcgctt 2001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 94.4  | 4.7         | 1288   | 14 | BQ678719    |
| C 2        | 90.6  | 4.5         | 1016   | 17 | CNS03L09    |
| C 3        | 88.2  | 4.4         | 1046   | 14 | BQ952554    |
| C 4        | 88    | 4.4         | 1651   | 12 | BG809816    |
| 5          | 86.6  | 4.3         | 1065   | 14 | BQ681076    |
| 6          | 86.6  | 4.3         | 1218   | 14 | BQ876717    |

|      |      |     |      |    |          |          |           |
|------|------|-----|------|----|----------|----------|-----------|
| 7    | 84.4 | 4.2 | 1339 | 13 | BM458211 | BM458211 | AGENCOURT |
| C 8  | 84.2 | 4.2 | 827  | 17 | AZ183795 | AZ183795 | SP_1002_A |
| 9    | 82.6 | 4.1 | 733  | 17 | AG039579 | AG039579 | Pan trogl |
| C 10 | 82   | 4.1 | 934  | 17 | AG046262 | AG046262 | Pan trogl |
| C 11 | 82   | 4.1 | 1023 | 17 | AG128304 | AG128304 | Pan trogl |
| C 12 | 81.6 | 4.1 | 1341 | 17 | AG030611 | AG030611 | Pan trogl |
| C 13 | 81.2 | 4.1 | 1242 | 14 | BM911414 | BM911414 | AGENCOURT |
| 14   | 81.2 | 4.1 | 1281 | 12 | BG852363 | BG852363 | 1024034A0 |
| 15   | 80.8 | 4.0 | 1051 | 14 | BQ689718 | BQ689718 | AGENCOURT |
| C 16 | 80.6 | 4.0 | 1116 | 12 | BG810038 | BG810038 | mgct002x1 |
| C 17 | 80.6 | 4.0 | 1233 | 17 | AG074706 | AG074706 | Pan trogl |
| C 18 | 80.4 | 4.0 | 1321 | 17 | AG126084 | AG126084 | Pan trogl |
| C 19 | 79.8 | 4.0 | 1798 | 17 | AG171124 | AG171124 | Pan trogl |
| C 20 | 79.4 | 4.0 | 1112 | 14 | BQ058195 | BQ058195 | AGENCOURT |
| C 21 | 78.2 | 3.9 | 666  | 12 | BG786337 | BG786337 | SEAU0006  |
| C 22 | 78.2 | 3.9 | 724  | 17 | AG171201 | AG171201 | Pan trogl |
| 23   | 78.2 | 3.9 | 1061 | 17 | AG046200 | AG046200 | Pan trogl |
| 24   | 78.2 | 3.9 | 1150 | 14 | BQ892481 | BQ892481 | AGENCOURT |
| 25   | 78   | 3.9 | 775  | 17 | AZ185516 | AZ185516 | SP_1005_A |
| 26   | 78   | 3.9 | 821  | 17 | AG063084 | AG063084 | Pan trogl |
| C 27 | 78   | 3.9 | 1243 | 14 | BM913931 | BM913931 | AGENCOURT |
| C 28 | 77.8 | 3.9 | 1143 | 10 | AW731158 | AW731158 | GA_Ea001  |
| C 29 | 77.6 | 3.9 | 715  | 17 | AG036665 | AG036665 | Pan trogl |
| C 30 | 77.6 | 3.9 | 1128 | 14 | BQ710735 | BQ710735 | AGENCOURT |
| C 31 | 77.6 | 3.9 | 1170 | 17 | AG111669 | AG111669 | Pan trogl |
| C 32 | 77.2 | 3.9 | 1297 | 14 | BQ064843 | BQ064843 | AGENCOURT |
| C 33 | 77   | 3.8 | 1025 | 17 | AG137560 | AG137560 | Pan trogl |
| C 34 | 77   | 3.8 | 1123 | 17 | AG080476 | AG080476 | Pan trogl |
| C 35 | 77   | 3.8 | 1134 | 17 | AG043044 | AG043044 | Pan trogl |
| C 36 | 77   | 3.8 | 1137 | 12 | BG809979 | BG809979 | mgct002xd |
| C 37 | 76.8 | 3.8 | 1268 | 13 | BM551006 | BM551006 | AGENCOURT |
| C 38 | 76.6 | 3.8 | 1127 | 14 | BQ718629 | BQ718629 | AGENCOURT |
| C 39 | 76.4 | 3.8 | 763  | 17 | AG083422 | AG083422 | Pan trogl |
| C 40 | 76.4 | 3.8 | 1197 | 13 | BI416470 | BI416470 | hasp001xa |
| C 41 | 76.2 | 3.8 | 1138 | 9  | AI374006 | AI374006 | SWOVAFCAP |
| C 42 | 76   | 3.8 | 1149 | 14 | BM910848 | BM910848 | AGENCOURT |
| C 43 | 76   | 3.8 | 1798 | 17 | AG171124 | AG171124 | Pan trogl |
| C 44 | 75.8 | 3.8 | 1250 | 13 | BM560215 | BM560215 | AGENCOURT |
| C 45 | 75.8 | 3.8 | 1424 | 14 | BQ720687 | BQ720687 | AGENCOURT |

ALIGNMENTS

|            |   |             |              |            |                 |
|------------|---|-------------|--------------|------------|-----------------|
| RESULT 1   | BQ678719  | 1288 bp     | mrna         | linear     | EST 15-JUL-2002 |
| LOCUS      | AGENCOURT_8199257   | NIH_MGC_112 | Homo sapiens | cdna clone | IMAGE:6261563   |
| DEFINITION | 5', mRNA sequence.  |             |              |            |                 |
| ACCESSION  | BQ678719  |             |              |            |                 |
| VERSION    | BQ678719.1  |             |              |            |                 |
| KEYWORDS   | EST.  |             |              |            |                 |
| SOURCE     | human.  |             |              |            |                 |
| ORGANISM   | Homo sapiens  |             |              |            |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |             |              |            |                 |
| AUTHORS    | NIH-MGC http://mgc.nci.nih.gov/.  |             |              |            |                 |
| TITLE      | 1 (bases 1 to 1288)   |             |              |            |                 |
| JOURNAL    | National Institutes of Health, Mammalian Gene Collection (MGC)  |             |              |            |                 |
| COMMENT    | Unpublished (1999)<br>Contact: Robert Strausberg, Ph.D.<br>Email: cgabs-r@mail.nih.gov<br>Tissue Procurement: DCTD/DTF<br>cdna Library Preparation: Rubin Laboratory<br>CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)<br>DNA Sequencing by: Agencourt Bioscience Corporation<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:<br>http://image.llnl.gov<br>Plate: LLCM2422 row: h column: 12<br>High quality sequence stop: 171.<br>Location/Qualifiers<br>1. :1288 |             |              |            |                 |
| FEATURES   | source  |             |              |            |                 |





```
Qy 281 CCCGACAGCTCTCTTACCGCTCCCGAACCAGGGTGGAGCTTCGTGCGCCGCCACCC 340
Dy 892 CCCGCCCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 833
Qy 341 GTCCACGAGAACACCATGAGCCCGGCATCTACGCGCGCGCGCTGCTCCAGGAG 400
Dy 832 CCGCGCCCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 773
Qy 401 CCGCGCCCGGAGCACTCGCTGATCTTCGCGGTGCTGTCGCGCTCGAGCCGGGCAT 460
Dy 772 CGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 713
Qy 461 CCGTCCCGGCTGCTCGTCCGCTTACCATATGACGCCGCCCGCCCGCGCTGCTCG 520
Dy 712 GCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 653
Qy 521 ACCCCACGCGCGCTACACACCGCGGCTACCACTGTACTACTGTACTGTCCGACC 580
Dy 652 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 593
Qy 581 AGAACACGCGCGCGCTGGGACACGAGACACGCGCTGCGCTTCACGC 640
Dy 592 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 533
Qy 641 ACCAGGACCGTATGCGGTGCGGCGCCGACTTCCCGCTGTGTCGCGGTGCGGTCG 700
Dy 532 GCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 473
Qy 701 TCGCACCGGACACGCGAGGTTCGCGCGCGCGGCGGTGCTGCGCTCGGACCGACG 760
Dy 472 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 413
Qy 761 CGACGCGCGCTCGCGAAGTACAGGAGCAGTACTTACTGTGTCGACGACGCGGT 820
Dy 412 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 353
Qy 821 TCAGTTTACCGCTGCGCGACCGCTGATGTTAAACACGAGGCTGCGCGCGCCACCA 880
Dy 352 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 293
Qy 881 CGCGCGCGGAGATGAGAACCGCGAGTGTTCGCGACCCCAAGATCCACTGGGACCG 940
Dy 292 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 233
Qy 941 CCGCGGAGATGCTGCTGCTATCGGACGACTCGGTACGCGCGCTTCTACACTCGC 1000
Dy 232 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 173
Qy 1001 CGACCTGCGGCTGAGACTTCCGCGCACTTCTGACTACCGAACACGCGCTCGGCG 1060
Dy 172 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 113
Qy 1061 GCATCGAGTGCCTGCTGTCAGATCAGCGAGACGAGGAGACGCGCAC 1113
Dy 112 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
```

```
RESULT 3
B0952554/c
LOCUS B0952554 1046 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8863775 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423858
5', mRNA sequence.
ACCESSION B0952554
VERSION B0952554.1 GI:22368032
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;
1 (bases 1 to 1046)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
```

```
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: DCTD/BTP/Gazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2603 row: b column: 19
High quality sequence stop: 206.
FEATURES
source location/Qualifiers
1..1046
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6423858"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

BASE COUNT 24 a 324 c 538 g 30 t 30 others

ORIGIN

Query Match 4.4%; Score 88.2; DB 14; Length 1046;  
Best Local Similarity 45.6%; Pred. No. 1.7e-05;  
Matches 379; Conservative 0; Mismatches 436; Indels 16; Gaps 3;

Qy 1 GCGGTGACCCCGACTTCCCTCGACGACACCGTCCCTTACGCGCGCGCGCGCCGCC 60  
Dy 988 GCG 929  
Qy 61 CGACTGCTCTCTAGCTAGAGCGCGCGCTTCTCGAGGTCTTCTGCGGAGCGGTGAGCG 120  
Dy 928 GCG 869  
Qy 121 ACTGCTGCAACCTGCTCTCTTGGGGCGCGCGGTGTGACCGGAGCTCGAGACGCA 180  
Dy 868 GCG 809  
Qy 181 CG----GCCAGGACCGTGCACGTGACCGCGTGCAGGCGCGCGCGCGCGCGCGCG 235  
Dy 808 GCG 749  
Qy 236 ACGCCCTTGAACCTGCGCGCGCTTCTGGGTGACGAGCGCTCCACCGCGACAGCTCT 295  
Dy 748 GCG 689  
Qy 296 TCTACCGCTGCGCAACCGAGGTGAGACCTTCTGTCGCGCGCGCGCGCGCGCGCG 355  
Dy 688 CCG 629  
Qy 356 CAGCAATGACCGCGCGCATCTACCGCGCGCGGTGCTTCCAGGAGCGCGCGCGCG 415  
Dy 628 CCG 569  
Qy 416 TCGCGCTGATCTTGGCGGCTGTGCGCGCTGACGCGCGCGCGCGCGCGCGCGCG 468  
Dy 568 CCG 509  
Qy 469 GCGTCTGCTGCGGTCTACCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528  
Dy 508 GCG 449  
Qy 529 GCGCGGTACACCGCGCGCGCTTACAGCTGTACTACTGTGCTGCTGCGCGCGCG 588  
Dy 448 CCG 393





|            |   |
|------------|---|
| RESULT     | 7   |
| BM458211   |   |
| LOCUS      | linear EST 05-FEB-2002  |
| DEFINITION | BM458211 1339 bp mRNA<br>AGENCOURT_6411296 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5530584 |
| ACCESSION  | 5', mRNA sequence.  |
| VERSION    | BM458211  |
| KEYWORDS   | BM458211.1 GI:18507251  |
| SOURCE     | EST.  |
| ORGANISM   | human.<br>Homo sapiens  |

```

SOURCE: Homo sapiens
ORGANISM: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE: 1 (bases 1 to 1339)
AUTHORS: NIH-MGC http://mgc.nci.nih.gov/
TITLE: National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL: Unpublished (1999)
COMMENT: Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12210 row: O column: 01
High quality sequence stop: 134.
FEATURES
    source      1..1339
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5530584"
                /clone_lib="NIH_MGC_71"
                /tissue_type="leiomyosarcoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo df.
                Average insert size 2.1 kb."
BASE COUNT      63 a 780 c 403 g 78 t 15 others
ORIGIN
Query Match      4.2%; Score 84.4; DB 13; Length 1339;
Best Local Similarity 47.9%; Pred. No. 7.4e-05;
Matches 414; Conservative 0; Mismatches 426; Indels 24; Gaps 5;
QY      2 CGGTGCACCCGCACTTCCCTCGACGACACACCGTCCCTCTACCGCGCGACGCCCGGCC 61
Db      111 1111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      253 CGCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
Db      111 1111 111 111 111 111 111 111 111 111 111 111 111 111
QY      62 GACTGTCTCTACGCTAGACGGGCGCCCTCTCTCGAGGTCTTGTGTGGGGAGCGGTGAGCGGA 121
Db      111 CGCCGCGCGCCCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 372
QY      122 CTGGGTGACCTGGTCTCTCTGGGGCGCGCGGTGTGACCGGAGCGCTCGACAGCGCAC 181
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      373 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 432
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      182 GGCAGGAACCGTGCACGTGACCGCATCGACGTGACGGCGCCAGCATGTGTAGCGCC 241
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      433 GCCCGCGCGCCCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 492
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      242 CTGAACGTGCCCGCGTCTGGGCTGAGGACGCTCCACCCCGACAGCTCTCTTCTACC 301
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      493 CGCGCGCGCGCGC-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 544
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      302 GCTGCCGGAACAGGTTGGACGCTTCTGTGCGGCCACCGCTCCACGAGAGGAACAGCAA 361
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      545 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 604
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      362 TGACGCGCGGCATCTCACGCGCGCGGTGTCTCCAGGAGCGCGCGGAGCACTCGCCC 421
Db      111 111 111 111 111 111 111 111 111 111 111 111 111 111

```









[illegible][illegible]

```
Query Match 4.1%; Score 81.2; DB 14; Length 1242;
Best Local Similarity 43.5%; Pred. No. 0.00025;
Matches 373; Conservative 0; Mismatches 480; Indels 5; Gaps 3;

Oy 43 CGGCGGACCGCGCGGCTGCTCTCAGCTAGAGGGCCCTCTCTCAGGTCCTC 102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 CCGGAGGAGTGGCGAGTGGCTGACCTGGACGATGATGGGGTCCCGCAGGCGTC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 103 GTCGGGCGGCTGAGGCGACTGCGTCGAACCTGGTCTCTCTGGGGCGGGGCTGTGACC 162
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 AGGAGGAGGNNNNNNNNNNCCNGNCCNGNNNGCCCGCGCCCGCCNGCCGCC 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 163 G-CGAGCTCGAGAGCGGACGCGCAGAACCGTGACGTGACCGGATGACGTGAGCG 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 GCGGCGCGCGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 222 GCCAGCGATGTCGACGCGCCCTGAACCTGCGCGCGTCTTGGGCTACGAGGCTCCAC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 282 CCGACAGCTCTCTTCTACCGTGTCCCGAACAGGCTGAGGCTTCTGTGCGGCCCGACCG 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 342 TCCAGGAGGAGAACAGCAATGACGCGCGGCAATCTACGCGCGCGCGGCTGTCCAGGAGC 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 402 CGGCGCGGAGCACTCGCGCTGATCTTCGCGGCTGCTGTGCGCGCTGACGCGCGGCGATC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 462 CGCTCCGGGCTCGCTTCTGCTGCTTACACATGACGCGCGCGCGCGCGCGCTGTGCGA 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 694 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 522 CCGCGAGCGCGGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 754 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 582 GAACAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT--TCAG 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 814 CNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 640 CACCAAGCGCGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 700 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 758 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 818 GGTTCAGTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 878 CCAGCGCGCGCGAGATCG 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1114 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
BG852363
LOCUS
DEFINITION
1024034A03.y2 C. reinhardtii CC-1690, normalized, Lambda zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION
BG852363
VERSION
BG852363.1 GI:14233547
KEYWORDS
EST.
```

```
SOURCE
ORGANISM Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE
1 (bases 1 to 1281)
Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 2
JOURNAL Unpublished (2000)
COMMENT
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1..1281
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with Exassist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 30 a 697 c 456 g 15 t 83 others
ORIGIN
Query Match 4.1%; Score 81.2; DB 12; Length 1281;
Best Local Similarity 43.1%; Pred. No. 0.00026;
Matches 418; Conservative 0; Mismatches 547; Indels 5; Gaps 3;
Oy 9 CCGGACTTCCTCGAGGACACCGTCCCTACCGCGGAGCGCGCGCGCGCGCGCTGCT 68
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 69 CCGGACTTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 129 GAACCTGTCTCTCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 189 AACCGTGACGTGACGCGGATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 249 TCGCGCGCGTCTTGGGCTGACGAGGCGTCCCGCGCGCGCGCGCGCGCGCGCGCG 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 309 GAACAGGCGGAGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 GCGCG--GGCGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 369 GGCCATCTCACGCGCGCGCGCGTCCAGGAGCGCGCGCGCGCGCGCGCGCGCG 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 672 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 429 CGCGGCTGCTGTCGCCCTGACAGCCCGGGATCCGCTCGGGGCTCGCTCGGCTCGCTCTA 488
Db 731 GGCCTCGGCGCCCGCCGCGGNCNCNGCCGCGGCGCCGCGGCGGCGCGCGCGCGCG 790
QY 489 CCACATGAGCCCGCCCGAGCGGCTGCTGCGAGCCCGCCAGCCCGCGGTACCACCCACG 548
Db 791 GCGGNCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 850
QY 549 GCGCTACCACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 608
Db 851 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
QY 609 GCGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 668
Db 911 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
QY 669 GCGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 728
Db 971 GNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
QY 729 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
Db 1031 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1088
QY 789 GCGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 848
Db 1089 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
QY 849 CATCGTCAACACGAGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
Db 1149 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1208
QY 909 GTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 968
Db 1209 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1268
QY 969 AGCAGTGGCG 978
Db 1269 CCGCAGCGG 1278

RESULT 15
BQ689718
LOCUS
DEFINITION BQ689718 1051 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
AGENCOURT_B344784 NIH_MGC_110 Homo sapiens cdna clone IMAGE:6247838
ACCESSION BQ689718
VERSION BQ689718.1 GI:21815034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1051)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC#2386 row: 1 column: 15
High quality sequence start: 2
High quality sequence stop: 283.
Location/Qualifiers
1. .1051
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
```

```
/clone="IMAGE:6247838"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pONB7; Site:1: XhoI;
Site:2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 63 a 557 c 291 g 69 t
ORIGIN
Query Match 4.0%; Score 80.8; DB 14; Length 1051;
Best Local Similarity 42.8%; Pred. No. 0.00029;
Matches 316; Conservative 0; Mismatches 415; Indels 8; Gaps 1;
QY 7 CACCCCGACTTCCCTCGACGACCGCTCCCTTACCAGCGCGCGCGCGCGCGCGCGCGCG 66
Db 310 CMCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 369
QY 67 CTCCTCAGCCTAGACGGGCGGCTCTCTCAGAGTCTTCTGTCGGGAGCGGTGAGGGAGTGG 126
Db 370 CCCCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 429
QY 127 TCGAACCTTGCTCTCTGCGGGCGGCGGTGTGACCGCGAGCTCGAGACGCGACCGGCA 186
Db 430 CCGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 489
QY 187 GGAACCGTGCAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
Db 490 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 549
QY 247 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 550 GCGCCNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
QY 307 CCGAACAGGCTGAGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 610 CCGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 569
QY 367 CCGGCGCTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 670 CCGCNCNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
QY 427 TTCGGCGGTGCTGTCGGCGCGCTCGAGCGCGG-----GGCATCCGCTCGGGGCTCGCTCC 478
Db 730 CNGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
QY 479 GTGCGGCTCTACACATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
Db 790 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 849
QY 539 CCACCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
Db 850 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 909
QY 599 GCTGGGACACGCGGAGCAGCGCGGTGCGCTTACGACACGCGCGCGCGCGCGCGCGCG 658
Db 910 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
QY 659 CGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 718
Db 970 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029
QY 719 CAGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
Db 1030 CGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1048
Search completed: May 10, 2003, 09:52:32
```

Job time : 2907 secs

---

**THIS PAGE BLANK (USPTO)**





```
QY 311 ACCAGGTGAGCTTCTGTCGGCCGCCACCCGTCACGAGGAGAACACGACATGAGCCCG 370
DB 3941372 GTGGCGCGCGGTGCTGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941313
QY 371 CCAATCTACGCGCGCGCTGCTCCAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 430
DB 3941312 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941253
QY 431 GGGGTGCTGTCGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490
DB 3941252 GTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941193
QY 491 ACATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
DB 3941192 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941133
QY 551 CCAACAGCTGTACTACCTGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 610
DB 3941132 CTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941073
QY 611 CGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 670
DB 3941072 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941013
QY 671 ACTTCCCGTGTGTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
DB 3941012 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940953
QY 731 CGGGCGCGGTGCTGCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790
DB 3940952 CGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940893
QY 791 AGTACCTTACTGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 850
DB 3940892 CTTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940839
QY 851 TGTCTAACACCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
DB 3940838 CGGTGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940779
QY 911 TCGGCGACCCAGATCCACTGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
DB 3940778 CGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940719
QY 971 GACTGCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
DB 3940718 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940659
QY 1031 ACTTCCGACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090
DB 3940658 CGCGCGCGCGGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940599
QY 1091 CGCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
DB 3940598 CGCGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940539
QY 1151 GCGTCCCGATGACGACGCGCTTACTGAGAGGACCGTGGAGCGCGCGCGCGCGCGCG 1210
DB 3940538 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940479
QY 1211 AGCACTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1270
DB 3940478 GCGCGTCCGCTGGCGCTGAAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940419
QY 1271 CATGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1330
DB 3940418 CTACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940359
QY 1331 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1390
DB 3940358 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940299
QY 1391 TCGGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1450
```

## RESULT 2

```
US-09-025-691-4
; Sequence 4, Application US/09025691
; Patent No. 6069299
; GENERAL INFORMATION:
; APPLICANT: Broadway, Roxanne M.
; APPLICANT: Harman, Gary E.
; TITLE OF INVENTION: FUNGUS AND INSECT CONTROL WITH
; TITLE OF INVENTION: CHITINOLYTIC ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,691
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/20120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
```

|    |      |  |      |
|----|------|--|------|
| Qy | 1348 | CCCACGAGCGATCCGACGGCTACACGGGCAGAACTCGATCGTCGGGAGCTCGGGTCC      | 1407 |
| Db | 1745 | GAGTCGGTGACGGGCTCGCGCACACTGGGACACGCCGTGGGGGAATTTCAACCAG        | 1804 |
| Qy | 1408 | GCCGACAGCCTGGCGCTGGTACACCCCTCTGAGCACCCCGTGGCAGCGCTGACGAAC      | 1466 |
| Db | 1805 | CTCCGCAAGCTCAAGGCCAAGTACCCGCGACATCAAGGTCTCTTGGTCTTTCGGCGGGCTGG | 1866 |
| Qy | 1468 | TACGTACACGCCCAACCAACATCCCCGACCGGACCGTGCAGGCGAGCGCC--GTCTTCG    | 1522 |
| Db | 1865 | ACCTGGTCCGGCGCTTACCGGACGCCGTGAAGAACC CGCGCCCTTCGSCCAAGTCTTCG   | 1922 |
| Qy | 1526 | CATGGAACGACGGCATAGAGAT-----CGAGTCGACATCGCTGG                   | 1568 |
| Db | 1925 | CACGACCTTGTTCGAGGACCCGCGCTGGGCGCAGCTCTTCGAGGGCATGACCTCGACTGG   | 1984 |
| Qy | 1570 | GACACCGCAGAACGTTCGGCATCTCGGTGGGCGCTCCCCCGACGGAACCCGGCACACG     | 1622 |
| Db | 1985 | GAGTACCCGAACGCTCGGCGCTCAGTGTGCAGAGCTCCGGTCCGGCCGCGCTGAAGAAC    | 2044 |
| Qy | 1630 | AACATCCGGCAAGTACGGAGCAGACCTGTAGCTGCACCGAGGACCTTCGACATCCGCCGG   | 1688 |
| Db | 2045 | ATGGTCAAGCGATGCGCGCCAGTTTCGGCAGGACCTGGTCACCGCCGCATACCGCC       | 2104 |
| Qy | 1690 | TACTGCTCGCCCTTACTCGGAGCGCGCCGCCCATCGACCCGGCGCCGATCCGTTG        | 1749 |
| Db | 2105 | GACGCCAGCTCCGGGCAAGCTCGACGCGCGGACTACGCGGGGCGCCCGCACTTC         | 2166 |
| Qy | 1750 | CACCTGGCGATCTCTGCACACCCAGAGGTTTCGAGTCTTCGTCACACCGCCGCACACC     | 1809 |
| Db | 2165 | GACTGGTACAACGTGATGAGTACGACTTCTTCGGCGCTTGGGCAAGACCGGCCCGACC     | 2222 |
| Qy | 1810 | GTGCTCTCCAGCAGGTCCAATTTCGCGGAGGGCGACACGGGAATCTCGCTCTFACACCGAC  | 1866 |
| Db | 2225 | CGGCCCACTGGCCCTGAATCTCTACAGGGGATCCCCAAGCGCGACTTCCATCTCGCC      | 2284 |
| Qy | 1870 | GGGGGCCCGCACACTTCAACGGCATCGTGTCCGGCAGA                         | 1909 |
| Db | 2285 | GCGCCATCCGCAAGCTCAAGGCGAAGGGCGTCCCGCGCA                        | 2324 |

135 GGTCCTCCTGGGGGCCGGGGGTGTGACCGCGAGCCTCGAGACGGCAGGCAACCGT 194

Db 3936308 CTTACACCGTGGCGCGTGGCTAGCCCGGTGCTTTAGCGCGCGCTTTGCGCGCTTGGC 3936249  
QY 195 GCACGTGACCGGATCGAGCTGAGCGGCCAGCGATGCTGACGCCCTGGAACCTGGCCG 254  
Db 3936248 CGCCGTGGCGCGCGCACCACTTGGCGCGCTCAATGCTGGGTCCCGTGGCGCGCG 3936189  
QY 255 CGTTCTGGGCTGACGAGCGCTGCC---ACCCGACAGCTCTCTTCTACCGTGGCCGAA 311  
Db 3936188 CGCCCGCGCGCGCGCGCGCGCGCGGTGCGCGCGATCCCGCGCTGCGCGGGAACACCG 3936129  
QY 312 CCAGGTGACGCTTCTGTCGCGCGCCACCGCTCTGACAGAGAGAACAGCAATACCGCGC 371  
Db 3936128 CGCGAATCCGGGTGTCGCGCGCGCGCGCGCGATGCGCGCTTGGCGCGCTTACCGCGCTT 3936069  
QY 372 CATCTACCGCGCGCTGCTCCAGGAGCGCGCGCGCGAGCACTCGCGCTGATCTTGG 431  
Db 3936068 GCGCGCGCGCGCGCGCTGCTGCGCACGCGCTGCTGCTTCTGGCGGTACCAACCG 3936009  
QY 432 CGGTGCTGTCGCGCTGACGCGCGCGCTGCGCTGCGCGCGCTGCTGCGCGCTGCTACCA 491  
Db 3936008 CCGCGCGGTGCGCGCGCGCGCGCGCGCTGCTGCGCGCGCGCGCGCTGCGCGCG 3935949  
QY 492 CATGACGCGCGCGCGCGCTGCTGCGACCGCGCGCGCTGCGCGCGCGCGCGCGCG 551  
Db 3935948 CGACCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935889  
QY 552 CTACAGCTGTACTACTGCTACGACGAGCAACAGCGCGCGCGCGCGCGCGCGCGCG 611  
Db 3935888 CGCTGTCG 3935829  
QY 612 GAGCAGCAGCGCGCGCGCTTACGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 671  
Db 3935828 CCGCGGTTCG 3935769  
QY 672 CTTCCCGGTGCTGCG 731  
Db 3935768 TGGCG 3935709  
QY 732 CGCGCGCGGTGCTGCG 791  
Db 3935708 CCGCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935649  
QY 792 GTACCTTACTGTCAGCGCGCGCGGTTCAGTTTACCGCGCGCGCGCGCGCGCGCG 851  
Db 3935648 CTTTCCCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935592  
QY 852 CGTCAACACCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 911  
Db 3935591 GGTCCCGTTCG 3935532  
QY 912 CCGCGACCGCAAGATCCATGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 971  
Db 3935531 CGTCGCGGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935472  
QY 972 ACTGGGTACCGCGCGGTTCATACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1031  
Db 3935471 GCGCGCGCGTCAACCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935412  
QY 1032 CTTGACCTACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1091  
Db 3935411 TCTGGCGGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935352  
QY 1092 CGCAGACCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1151  
Db 3935351 CGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935292  
QY 1152 CTTCCCGATGACGTACCGCTTACTGACAGGACGCGCGCGCGCGCGCGCGCGCG 1211  
Db 3935291 CGTCTTGGGAGAGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3935232  
QY 1212 CGACCTACCGCGCAATGGCTGACTGGGTGGGACTGGTACGCGCGCGCGCGCGCG 1271

Db 3935231 CGCGATATACCGCGCTTGGCGCGCGGTTCGCGCGCGGTTCGCGCGCGCGCGCG 3935172  
QY 1272 ATCGATGACGCGCGCGCGGACCAAGCGCTCGCCATCGCGTGGATGAACAACCTGGA 1331  
Db 3935171 GCGCGCGGTTCG 3935112  
QY 1332 CG 1391  
Db 3935111 TGGCG 3935071  
QY 1392 CGCGGAGTTCG 1451  
Db 3935070 CTGCGCGCGGTGGCTAGCGCGCGGTGCTGTTAGCGCGCGCTTTCGCGCGCT 3935013  
QY 1452 GCGAGCGTTCGAGAACTAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1511  
Db 3935012 GCG 3934953  
QY 1512 CAGCGCGGTTCG 1571  
Db 3934952 CCG 3934905  
QY 1572 CACCG 1631  
Db 3934904 CACCG 3934845  
QY 1632 CATCG 1691  
Db 3934844 ATCG 3934788  
QY 1692 CTCGCTGCG 1751  
Db 3934787 GTCG 3934728  
QY 1752 CCG 1811  
Db 3934727 CG 3934668  
QY 1812 GCTCTCCCG 1871  
Db 3934667 GCGCGCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934608  
QY 1872 CG 1931  
Db 3934607 CG 3934548  
QY 1932 GCACACCG 1962  
Db 3934547 CACCG 3934517

RESULT 4

US-08-403-852D-1  
; Sequence 1, Application US/08403852D  
; Patent No. 5891695  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanche, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Laurent  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; Biosynthesis Of Streptogramins, Nucleotide Sequences  
; OF THESE POLYPEPTIDES AND THEIR USE  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5392 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
US-08-403-852D-1

Query Match 4.6%; Score 91.4; DB 2; Length 5392;  
Best Local Similarity 42.1%; Pred. No. 1.8e-08;  
Matches 652; Conservative 0; Mismatches 891; Indels 6; Gaps 2;

Qy 382 CGCGCCGTGCTCCAGGAGCGCGCGGAGCACTCGCCCTGATCTTCGGCGGTGCTGTG 441  
Db 2748 CGGCACGATACACCGCGCGCTGCGGGCGCGGAGCGCGCGGATTCGACGCGCTCTGT 2807  
Qy 442 CCGCTGCGAGCGCGGATCGCTCGGGCTCGCTCCGCTGCTTACACATGACGCGCC 501  
Db 2808 CGAGCAGCGCGCGCGCGCTCCAGGCGGTTTCGAGACGAGAC--CTGACCGCC 2865  
Qy 502 CCCAGCGCTGGCTCTGCGACCCCAAGCCCGGTTCACCCACCGCGGCTTACAGCTG 561  
Db 2866 GCGTGGCGCGCTGACCGAGCACATCGGCTGATACCGCGCGCTCCCGCGCGACAG 2925  
Qy 562 TACTACCTGCACTCCGACCAACAGCGCGCGCGCGCTGGGACCGGAGCAGCAGCC 621  
Db 2926 GCCCTTACCAAGCTGTCGGATCACCGCTCGCTCGACCACTCGGCCACCGCGCGC 2985  
Qy 622 GACGCGCTGCGCTTACGACCAACAGCGCGCGCTGATGCGCGCTGCGCGCGGCTTCC 681  
Db 2986 GGCTGGCTCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3045  
Qy 682 TGGTCCGGGTGCGGCTGCTGCGACCGCGAAGCAGCGGAGGTTTCGGCGCGCGCGCTG 741  
Db 3046 GTCGTCCGCGCTGTGGAGAGCTTCGACGACGAGCGCTTCGTCACGACCGCGCGCGC 3105  
Qy 742 GTCGGCTCGGACCGCAGCAGCGCGCTTCGCGAGTACGAGGAGCAGTACCTCTAC 801  
Db 3106 GGCCTGTACTGGCGCTGCGCGCGCTTCACCACTCGACCAAGGCGAGGAGCAGTTCGAC 3165  
Qy 802 TGGTGCAGCGGCGGTTTCAGCTTCACCGCTTCACCGCGCGCGCGCTGCTGCTGCTG 861  
Db 3166 GTGCGCGCGCGCTTACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3225

Qy 862 GACGCTCGCGCGCGCACCGCGCGCGGAGATCGAGAACGCGGAGTGTTCGCGACGCC 921  
Db 3226 GCGCGCGCGCTCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCGAGCGCGCGCGCGCG 3285  
Qy 922 AAGATCCACTGGGACACCGCGCGCGGAGAAATGGGTCTCGCTCATCGGACGACTGGGTAC 981  
Db 3286 TCGGTGAAGCAGCAGCG 3345  
Qy 982 GCGCGCTTCTACACTCGCTCG 1041  
Db 3346 GAATGCG 3405  
Qy 1042 CCGAACACCGCGCTCG 1101  
Db 3406 CCGGTCTTGGCGCGCTCG 3465  
Qy 1102 GGGACACCGCGCTCG 1161  
Db 3466 CTGCGCGAAGCGCTGGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3525  
Qy 1162 ACGTACGCTTACTGGACAGCG 1221  
Db 3526 CGACCGCTCG 3581  
Qy 1222 CCGCAATGGCTCGACTGGGCTGGGACTGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1281  
Db 3582 GAGTCCGTGACCGGCG 3641  
Qy 1282 GCGCGCGAGACCAAGCGCTCGCGCTCGCGCTGGATGAACAACACTGGAAGTACCGCGCGCG 1341  
Db 3642 GACGCGCTGCTGCGCGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3701  
Qy 1342 GACGTCCCG 1401  
Db 3702 GCGCAGGTCCACATCG 3761  
Qy 1402 CGGCTCG 1461  
Db 3762 GTCCGCGACGAGATCTCGCGCTCGCGCTGACGACTCGTCCGCGCGCGCGCGCGCGCGCG 3821  
Qy 1462 ACGAACTACGTTCACCG 1521  
Db 3822 TCCTGCGCGCTGCTCGCTCTCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3881  
Qy 1522 CTGCATGGAAGCG 1581  
Db 3882 AGCGCTACGAGACCGCGCTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941  
Qy 1582 AACGTGCGCATCTCGGTGGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1641  
Db 3942 GACGAGGCGCTGATGTTGCGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4001  
Qy 1642 TACGAGCAGACCTGTAGTTCG 1701  
Db 4002 ATCAGGTGCG 4061  
Qy 1702 CCCTACTCG 1761  
Db 4062 CCCTACTCG 4121  
Qy 1762 CTCGTGACACCG 1821  
Db 4122 GTGGCGCTGGACACCGCTGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4181  
Qy 1822 CAGGTCCACTTCG 1881  
Db 4182 CTGCTACCG 4241  
Qy 1882 CACTTTCACCG 1930  
Db 4242 GACGCGCATCAAGCTCGAGACGCGCAACTACCGCGCTGCTGGTCAACCGCA 4290











Db 26238 CCGACACCCGGTTCGACGACGGGGCCCGCTGCTGCGCGAGAACTTCACGGAGCTGATCGC 26297  
QY 129 GAACCTGGTCTCTCTGGGGCCGGCGGTGTGACCGGAGACCTTCGAGACGGCAGCGG 188  
Db 26298 GGGCCTCGACGACCTCTGTCGAGGGCCGACCGCACCGCTCTGCTGCGGGGACCGCCGG 26357  
QY 189 AACCGTGCAGTGACCGCGATGACGCTGAGGGGCGCCACGCGATGTGACGCCCTGAACC 248  
Db 26358 CACCTCCGACAGTGTGCTGCTCTCCCGGCCAGGGCTCGCAGTGGCCCGAGATGC 26417  
QY 249 TCGCGCGTCTGCGGTGACGAGCGCTCCACCCGACAGCTCTCTTCTACCGTGGCC 308  
Db 26418 CGAGGGCTCTGGCCGCTCCAGCGGCTCCGGCTCTCTTCTTGAGACCGCCCGCGCTG 26477  
QY 309 GAACAGGGTGGACGCTTCTGTCGCGCCACCGCTCCACGAGAGGACCAATGACGCG 368  
Db 26478 CGACCTCGCGCTCGGCCCCACTCTGGCTGGTTCGCTCTGAGAGTACTCGCGGGGAACC 26537  
QY 369 GGCATCTCACCGCGCGCGTGTCTCAGGAGCGCGCGGAGCACTCGCCCTGATCTT 428  
Db 26538 CGCGCGCGCTCTGCTGACCGGCTGACGCTGTGACGCGCTGCTTTCACCATGATGT 26597  
QY 429 CGCGGTGTCTGCGCGCTGACGCGCGGCTGCGCTCGCGGCTGCTCGCTGCGCTCTA 488  
Db 26598 CTGCTCGCGGAGACGTGGCGTTCGCTGGCGCTGGAACCGCGCGGCTGCTCGGTACATC 26657  
QY 489 CCACATGACCGCCCGCGGCTGCTCTGACACCGCCCAACGCGCGGTCAACACCCACCG 548  
Db 26658 CCAGGCGAGATCGCGCGCGCTACGCTCGCGCGCGCTGACGCTGAGACGACGCGCG 26717  
QY 549 CGCCTACGAGCTGTACTCTGACTCTGACGAGGAGCGCGGAGCAACGCGCGCGGCTGGAGCA 608  
Db 26718 CATCTGCGCGCTGCGAGCAGCGCTGGCTGGCGGCTGCGCGGAGCGGCGGATGCTCG 26777  
QY 609 CGCGACGACGACGCGCGGTGCGCTTTCACGACGACGCGACCGGTGATCGCGTGGCG 668  
Db 26778 CGTGACCTGTGCGAGCGGACCTGCGTCCCGCTGAGCGCTGAGGAGCGCGCTCGC 26837  
QY 669 CGACTTCCCGTGTGCTGCGGTGCGCGTCTGCGGACCGGAGCAACGCGGAGGTT--- 725  
Db 26838 CGTGGCGCGCTCAACGGCGCGGAGACCTGCGCGCTCTCCGGGACCGCGCGCTGCG 26897  
QY 726 -----CGGGCGCGCGGCTGCTGCGGCTGCGGCGCGGCGGCGGCTGCGGCTGCGGCT 773  
Db 26898 GGAGCTGTGCGCGAATCTGCTGCGGAGGCGGTGACGCGCGCGCCCATCCCGGCGTCA 26957  
QY 774 CCGAAGTACGAGGAGCTACTCTACTGTGTCGACGCGCGGCTTCACTGTCACGCG 833  
Db 26958 CACCGCGGCGACTCGCGCGAGTTCGACAGCTGAGGCGCCACCTGCGGAAGTCTCGC 27017  
QY 834 CTGCGCGCGCGTATCTCAACACCGACGCTGCGCGCGCGCGCGCGCGCGCGCGAGAT 893  
Db 27018 GCGCGTGGCGCGCGCGCTCGACATCCGCTTCTACTCGAGCTACCGGCTACCGGAGGATGAT 27077  
QY 894 CGAGAACCGGAGTGTTCGCGGACCGCGAAGATCTGAGACCGCGCGCGCGGAGATG 953  
Db 27078 CGACACCGCGAGCTGGAGCGCGGCTACTGCTGACGACATGCGGAGCGCGTGGAGTT 27137  
QY 954 GGTCTGCTCATCGGACGACTCGGTACCGCGCTTCTACCTCGCGCAACTCGCGCA 1013  
Db 27138 CGAGCGGCGCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 27197  
QY 1014 CTGGAGACTTCGCGCGAATCTGACTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 1073  
Db 27198 GCACCCCATGCTGGCGTCTCCCTCCAGGAGACGATCAGCGCGCGGTTCGCCGCGCG 27257  
QY 1074 CGACCTGTCGAGATCACCGGAGACGCGGACCGCGCGCGCGCGCGCGCGCGCGCGAT 1133  
Db 27258 CGTCTCTCGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTG 27317  
QY 1134 GGACGCGCTACCGCATCTCCCGATGCTACGCTACTGAGACGAGCGCGCGCGCGCGCG 1193  
Db 27318 CCGCGCGCTACACCGCGCGCTGGAGATCGACCGCGCGCGCGCGCGCGCGCGCGCGCTCAGC 27377

QY 1194 CGAGCAGTTCCAGCGCGACGACCTCACCCGCAATGGCTCGACTGGGGCTGGGACTGGTA 1253  
Db 27378 CCAGGTGGAATGCCACAGTACCCCTTCAGCGCGAGCGCTACTGTGTACA:CCCCGGCCA 27437  
QY 1254 CCGCGCGCTACCTGGCCATGATGACGCGCGCGCGGAGACCAAGCGCTCGCCATCGCGTG 1313  
Db 27438 CCGCGGTGACGACCGCGCTCTGCTGAGCGCGCTGACCGCGCGCTGCGGCGAG 27497  
QY 1314 GATGAACAACTGGAAGTACGCGCGACGCGCTCCCGACCGGACGATCGCGCGCTACAA 1373  
Db 27498 CCGGTGGAATGCGGAGTCCGCTGACCGGATGTACCGCGCGCGCTGGCGCGCGACAC 27557  
QY 1374 CCGGCGAGAACTCGATCTCGCGGAGTCCGCTCGCGCGCGCGCTGGCGGTGGTGTACAC 1433  
Db 27558 CACCGGTGCTGGCGGACCGCGCTGCTGGGTGCGCGCTGCTGCGCGCGCGCTT 27617  
QY 1434 CTCTCTGACACCGCGTGGCGAGCGTACGAACTACGTACGCGCGCGCGCGCGCGCGCG 1493  
Db 27618 CG-----CCGACCTGGCGCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 27668  
QY 1494 CGACCGGACCGTCCGCGCGCGCTGCTGCTGCAATGGAACGCGCGCGCGCGCGCGCG 1553  
Db 27669 GGAGTCACTTGGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 27728  
QY 1554 GCTCGACATCGCTGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1613  
Db 27729 GAACGTGGCG 27788  
QY 1614 CGGACCG 1673  
Db 27789 GGGCG 27842  
QY 1674 CTCCGACCTCGCGCGCTACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1733  
Db 27843 GCG 27902  
QY 1734 CCG 1779  
Db 27903 CCGCGCGCTTCTACGCGCGCTTCTCGGAACTCGGTTACGCGTACGCG 27948

## RESULT 9

US-07-945-283-1/c  
; Sequence 1, Application US/07945283  
; Patent No. 5352596  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Andrew K.  
; APPLICANT: Wesley, Ronald D.  
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
; TITLE OF INVENTION: Involving The EP0 and LLT Genes  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis P. Ribando  
; STREET: 1815 No. 5352596th University Street  
; CITY: Peoria  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 61604  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/945,283  
; FILING DATE: 19920911  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ribando, Curtis P  
; REGISTRATION NUMBER: 27976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 309-685-4011 ext.513

```
TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; NAME/KEY: variation
; LOCATION: replace(1267, "t")
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
;
; US-07-945-283-1
Query Match 4.2%; Score 84; DB 1; Length 8438;
Best Local Similarity 43.0%; Pred. No. 3.7e-07;
Matches 640; Conservative 0; Mismatches 830; Indels 18; Gaps 4;

Qy 329 TCGGCGCCACCGCTCCAGAGAGAACCAAGCAATGACGCCGCCATCTACGCCCGCGCG 388
Dy 5549 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5490
Qy 389 TGCTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448
Dy 5489 CTCTCGCGCTCCGCCACCTCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5430
Qy 449 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508
Dy 5429 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5370
Qy 509 GCTGGCTCTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
Dy 5369 ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5310
Qy 569 TGCACCTCGACAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628
Dy 5309 CCGACCGCTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5250
Qy 629 TCGCGCTTCACGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 688
Dy 5249 TCTGGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5190
Qy 689 GGTGGCGCGGTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 748
Dy 5189 TGCGCGCTTCATCCGCGAGATGGGGACTCGAGGAAGCAGCAGCGCGCGCGCGCGCG 5130
Qy 749 TCGCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 808
Dy 5129 TCATCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5070
Qy 809 CCGAGCGCGGTTCACGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868
Dy 5069 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5010
Qy 869 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 928
```

## RESULT 10

```
US-08-951-742-1
; Sequence 1, Application US/08951742
; Patent No. 6127144
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Michael J. Butler
; APPLICANT: Dany Hadary
; APPLICANT: David Jenish
```



RESULT 11  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 4.1%; Score 83; DB 4; Length 4403765;  
Best Local Similarity 42.6%; Pred. No. 6.6e-07;  
Matches 774; Conservative 1; Mismatches 1029; Indels 12; Gaps 4;

QY 86 CCCTCCTCGAGGCTCTCGTCGGGGAGCGGTGAGCGACCTGCGTCGAACCTGGTCCTCTGG 145  
DB 1631220 CCCTCGTGGTCGAGAGACGCCCTTGGCGCGGTGCTGCGCGCGCGCCAGTCCCGCG 16312179

QY 146 GGGCGCGGGTGTACCGGAGCCTTCGAGACGGCAGCGCCAGAACCTGTGACGTGACCG 205  
DB 1631280 ATGCCACCGTTGCCCGCGTGGCGCGGTGCGCGCTTACCGCGCTTACCCCGCTGGCG 1631339

QY 206 CGATCGAGCTGAGCGCGCCAGCGATGTGAGCGCCCTTGAACCTGCGCGCTTCTGGGCT 265  
DB 1631340 CCGGGCGCGCGTTCGCGCGCTGGCAGCGCGGTGGCGCGGTGACCGCGCTTCCCGCG 1631399

QY 266 GACGAGCGCTCCACCCCGACAGCTCTCTCTACCGCTGCGCGACACAGGGGTGACGCT 325  
DB 1631400 TCGTGCMGGATGCTCCGCGCGCGCCCGCTTGGCTGCGGTGGAGCGCGTCCCGCGGG 1631459

QY 326 TCGTCGCGCCACCGCTCCAGAGAGAACAGCAATGACGCGCGCCATCTACGCGCGCG 385  
DB 1631460 CCGCGCGCACCGCGTTCGCGCGCTTGGCGCGCTTGGCGCGCTGCGCGCGCGCGCG 1631519

QY 386 CGGTCTCCAGGAGCGCGCGCGAGCACTCGCCCTGATCTTCGGCGTGTGCGCG 445  
DB 1631520 GCGAAGCGCCCTGCTCCCTGCGCGCTTGGCGCGCTTGGCGCGCTTGGCGCGCGCGCG 1631579

QY 446 CTGACGCGCGGATCCCGCTCGCGGCTCGCTCGCGCTTACCACTATGACGCGCGCCCA 505  
DB 1631580 TCGCGCGCGCGCCACCGGACCGCGCGCGCTGCTGACTAGCGCTTACCGCGCGGTG 1631639

QY 506 GGGGTGCTGTGACCGCCCAACCGCGGTGACACCGCGCGCTTACCGAGTGTACT 565  
DB 1631640 CCGCGCGCGCGCGTTCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGGTGTGCG 1631699

QY 566 ACCTGCACTCCGACAGAACAGCGCGCGCGGTGGGACCAACCGGAGCAGCGCGAGG 625  
DB 1631700 GCGGACCGGACCGCGCTTGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631756

QY 626 GCGTGCCTTACGACCAACCGCGCTGATGCGCGCTGCGCGCGCTTCCCGCTGTGCT 685  
DB 1631757 AGCGCGCGTTCGCGCGCGCGCGCGCGCTTACCGCGCTTACCGCGCGCGCGCGCGCG 1631816

QY 686 CCGGTGCGCGTGTGTCGGCACCGCGAACAGGCGAGGCTTGGCGCGCGCGCGCGCGTGTG 745

DB 1631817 CCGCGCGCAC---CGTTGCCCGCGGGTTACCGTTTGGCCCGTTTACCAAGGTTGGTG 1631873

QY 746 CGCTCGGACCGGACCGGAGCGGTCGCGCAAGTACGAGGAGTACTCTACTGTT 805

DB 1631874 GCGTTGGCACTCATGCCACCAAGCGCGCTCGCGCGCGCGCGCGCTTCCCGCGCGTG 1631933

QY 806 CGACCGAGCGGGGTTACGTTTACCGCCTGCCGACCGCGCTCATCTGCTCAACACCGAGG 865

DB 1631934 CCGGCGCTGCGCGGTTGCGCCCATTTGCCCGCTGCGCGCGCTTCCCGCGCGCACTGG 1631993

QY 866 GTGCGCGCGCCACCGCGCGCGAGATCGAAGACGCGAGTGGTTCGCGCGACCGCAAGA 925

DB 1631994 GAGTTGCGCGGTTGCGCGCGCTGCGCGCGCTGCTGGAGTGAAGCGCGTG 1632053

QY 926 TCCACTGGGACACCGCGCGCGAGATGGGTCTGCTATCGGACGACTGCGGTAGCGCG 985

DB 1632054 GCGCGCTTGGCGCTGGGTTAGAGCGCGCGCTTACCGCTGCGCGCGCGCGCGCG 1632113

QY 986 CGTTCTACACCTTCGCGCAACCTGCGCGACTGGACACTTCGC---CGCAACTTCGACTACC 1042

DB 1632114 GGGTTACCGCGTTTACCGCGGTGACCGCGTTACCATCGCGAAGGCGAGTTGCCGTG 1632173

QY 1043 CGAACACCGCTCGCGCGCATCGAGTGCCTGCGCGACTGTTTCGAGATCAACCGAGAGG 1102

DB 1632174 GCGCGCTTGGCGCGCTCACCGCGAGCGCGCGCGCGCGCGCTTTCGCGCGCGCGCG 1632233

QY 1103 GGACAGCGCACTGGGTGCTGCGCGCGCGCATGGAGCGCTACGGCATCGCGCTCCCATGA 1162

DB 1632234 ACACCTTGATTCGCTTCTGGCGCAAGAGGTTCCCGCGCAACCGCGCGCGCGCTTGG 1632293

QY 1163 CTTAGCGCTTCTGGAGCGCACTTGGGACCGCGCGCGAGTTCACCGCGCGCGCGCGCG 1222

DB 1632294 CCGCGCTTACCGCTTGGCGCGCGCGCGCGCTGCGCGCTGCGCGCTTTCGCGCG 1632353

QY 1223 GCAATGCTGACTGGGGCTGGGACTGTAAGCGCGCGCTACCTGCGCGCATGATCGAGG 1282

DB 1632354 GCGCTGTTGGGCTTGGCGCGCTTGGCGCGCTTGGCGCGCGCGCGCGCGCTTGG 1632413

QY 1283 CGCGGAGACCAAGCGCTCGCATCGGTGGATGAACACTGGAAGTACSCCGCGCGCG 1342

DB 1632414 CCGCGCGCGCGCGCTGCGCGCGCTGCG---CGCGAAGTTCGCGCGCGCGCGCGCG 1632470

QY 1343 AGTCCCCACGACGATCCGACGCGTACACGGCGAGAACTCGATCTGCGCGCGAGCTGC 1402

DB 1632471 CCGCGCGCGCGCGCGCGCGCTTTCGCGCGCTGCGCGGATTCGCGCGCGGTG 1632530

QY 1403 GGTGCGCGGACGCTGGGCTGTGTAACCGCTCCTGAGCACCGCGCTGGGAGCGCTGA 1462

DB 1632531 GCGAGGAACCGCGCGACACCGCGCATGCGCTTCCCGCTTTCGCGCGCGCGCGCG 1632590

QY 1463 CGAAGTACGTCACCG 1522

DB 1632591 CCAACAGCGCGCGCTGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632650

QY 1523 TGCCATGGAAGGAGCGCATACGAGATCGAGCTCGACATCGCTTGGGACACCGCGAG 1582

DB 1632651 GCGTTCGCGCGTTCGCGCGCGCTGCGCGCGCTTTCGCGCGCGCGCGCGCGCG 1632710

QY 1583 AGTGGCGATCTGGTGGCGCGCTTCCCGCGAGGAACCGCGCGCGCGCGCGCGCGCGCG 1642

DB 1632711 GCACCGTCTGACCGCGCGGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632770

QY 1643 ACGGAGGACGCTGTACCTGCGACGAGGACCTCGGACTCGCGCGGTGCTGCTGCGCG 1702

DB 1632771 GCGTACCGCGGTTCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632830

QY 1703 CTTACTCGGAGCG 1762

DB 1632831 TTGCGACCGTTCG 1632890

QY 1763 TGTGCGACACCGAGCGCTGAGGCTTTCGTAACCGCGCGCGCGCGCGCGCGCGCGCG 1822



|    |      |   |      |
|----|------|---|------|
| Qy | 1690 | TACTCGCTCGCCCTCTACTCGGAGCGCGCGCCGCCATCGACACCGGGCGCCGATCGGT    | 1749 |
| Db | 1166 | GAGCTGATCTCTTTCTTTCAGAGCGCCCGTACTACGACTCCGCTGGGCGCGACCGCG     | 1225 |
| Qy | 1750 | CACCTCGGATCTCGTCGACACCCAGAGCGTCGAGGTCTTCGTCAAGCGCGGCCACACC    | 1809 |
| Db | 1226 | GAGATCTTCAGCAAGTACGTCGCGCGGCGACACCCAGGCGCTCGTCGACGCGCGCGCACCC | 1285 |
| Qy | 1810 | GTGCTCTCCAGCAGTCCACTTCCGCGAGGCGGACACGGGAATCTCGCTCTACACCGAC    | 1869 |
| Db | 1286 | GACCTGTCCGACACCGGGCAACGCTCCGGGAGAACGGCAACGCCGCTCTACACGGCC     | 1345 |
| Qy | 1870 | GGCGGCCCGCACACTTCACGGCATCGTCGTCGG                             | 1904 |
| Db | 1346 | GTGAGCTGCACGCGCCCAAGTGGCGCCGCAACTG                            | 1380 |

## RESULT 13

```

US-08-265-310-1
; Sequence 1, Application US/08265310
; Patent No. 5856166
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Butler, Dany
; APPLICANT: Hadary, David
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; APPLICANT: Malek, Lawrence T.
; APPLICANT: Soostmeyer, Gisela
; APPLICANT: Walczyk, Eva
; APPLICANT: Krygsman, Phyllis
; APPLICANT: Garven, Shelia
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/265,310
; FILING DATE: 24-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,508
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/133/CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1908 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..1759
; FEATURE:
; NAME/KEY: misc feature

```



Db 926 CAGCTCGCGCGCATGTGCTGCACAGGTGCTCAACCCGTCCTCGCGACAGATCTCGTAC 985  
QY 1510 GGACGCGCGCTCTGCTGCAAGACGCGCATACAGATCAGCTCGACATCGCTGG 1569  
Db 986 CAGGCCAAGCTGACAGGAGCTGCGCTTCGAGGGCGCTGGAAGACTGGAGGACTGG 1045  
QY 1570 GACACGCGGACGAACTGCGCATCTCGGTGGGCGCTCCCGCGGAAACCCCGCACAG 1629  
Db 1046 GTGCGCGGAAACGACGCGCTACCACTCGCGGACACCGCGCGAGGTCCAGGACCAG 1105  
QY 1630 AACATCGGAAGTACGAGCAGACCTGTAGCTCGAGCGGAGGACCTCCGACCTCGCGGG 1689  
Db 1106 TGCGTGAAGCTGCGCGCGCGCGCGGCGGAGAGCCGCTGGGGCGCTCGTGGACCGGG 1165  
QY 1690 TACTCGCTCGCCCTTACTCGGAGCGCGCGCGCCCATCGACCCCGCGCGCATCCGTG 1749  
Db 1166 GAGCTGATCTCTTCTTCAGAGCGCGCGCTACTAGACTCCGCTGGCGCGCGCGG 1225  
QY 1750 CACCTCGGATCTCTGTCGACACCCAGAGGTGCGGTGTCATGTCACCGCGCGCACAC 1809  
Db 1226 GAGATCTTACGAAGTACGTGCGCGCGGACACCCAGGCGCTGTCGACGCGCGCGCAC 1285  
QY 1810 GTGCTCTCCAGCAGTCCACTTCGCGGAGGCGGACACGGGAATCTCGCTTACACCGAC 1869  
Db 1286 GACCTGTCGACACCGCGGCGGCAACGCTCCGCGGAGAACGCGCAACCGCTTACACGG 1345  
QY 1870 GCGCGCGCGCACATTCACCGGCATCTGTCGCGG 1904  
Db 1346 GTCAGTGCACGACGCGCAAGTGGCGCGCGCAACTG 1380

## RESULT 14

US-08-387-942C-1  
; Sequence 1, Application US/08387942C  
; Patent No. 5939289  
; GENERAL INFORMATION:  
; APPLICANT: ENTERVAG, HELGA  
; APPLICANT: VALLA, SVEIN  
; APPLICANT: SKJAK-BRAEK, GUDMUND  
; APPLICANT: LARSEN, BJORN  
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES  
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/387,942C  
; FILING DATE: 09-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1809-106P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12588 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)

## ORIGINAL SOURCE:

ORGANISM: Azotobacter vinelandii

STRAIN: E

FEATURE:

NAME/KEY: CDS

LOCATION: 290...1951

FEATURE:

NAME/KEY: CDS

LOCATION: 2227..6438

FEATURE:

NAME/KEY: CDS

LOCATION: 6702..9695

FEATURE:

NAME/KEY: CDS

LOCATION: 9973..12588

US-08-387-942C-1

Query Match 4.0%; Score 80.4; DB 2; Length 12588;

Best Local Similarity 42.3%; Pred. No. 1.7e-06;

Matches 508; Conservative 0; Mismatches 691; Indels 3; Gaps 1;

QY 705 CACCGGAAACACGGCAGGTTCGGCGCGCGGCTGCTCGCTCGCGACCGCGGAC 764

Db 6875 CATCAAGACGACGTCATATCGTCGGCGGGGATGGGGAGACGCTCATCAAGTGGT 6934

QY 765 CGACGCGCTCCGCAAGTACGAGGAGTACTCTACTGTCGACCGAGCGGGGTTCAC 824

Db 6935 CGACGCGCTGGGATCAGGACGTCACCGCATCGTCCGCTACGGCGAGGAGACCAG 6994

QY 825 GTTACCGCGCTCGCGGACCGCGCTATCGTCAACACCGAGGTCGCGCGCGCACAGGCC 884

Db 6995 CAACCTTCGGCATGAGCGACCTGACCTTCGACGGCAACCGCACACGCGGCAAGGT 7054

QY 885 CGCCGAGATCGAGACGCGGAGTGGTTCGCGGACCCCAAGATCCACTGGGACACCGCCG 944

Db 7055 CGACGCGCTGTTCAACGGTACATTCGCCGCGAGGAGCGGCGGCGGACGTGACCT 7114

QY 945 CG---GAGAATGGTCTGCTCATCGGACGACTGCGGTACGCGGCTTCTACACCTCGCC 1001

Db 7115 GGAGCGGTGGAATCCGTGAAATGTCGGTTACGGTTTCGATCCGACGAGCAGACCAT 7174

QY 1002 GAACCTGCGGACTGACACTTCGCGGCACTTCGACTACCGAACCGCCCTCGGCGG 1061

Db 7175 CAACCTGACGATCCGCGACGCGTGGCCACGACAACGCGCTCGACGGCTTCGCGCGGA 7234

QY 1062 CATCGAGTCCCGGACCTGTCGAGATCACCGCAGACGCGGACACGCGACTGGGTGCT 1121

Db 7235 TTTCCAGATCGCGGGGTGTTTCGAGAACACGCTCTCGTACAACAGCAGCGCGCTT 7294

QY 1122 CGCGCGCAGATGGAGCGCTACGGCATCGGCTCCCGCTAGCGTACGCTACTGGACAGG 1181

Db 7295 CAACATCGTCACGACGACCAAGACTTCGCTCTGAGCAACAACGCTGCGCTACGGCAACGG 7354

QY 1182 CACCTGGGACGCGGACGAGTTCACCGCGCAGCAGCTCACCCCGCAATGCTCGACTGGG 1241

Db 7355 CGCGCGCGCGCTGGTGGTGCAGCGCGCTGTCGCGACGTGGCGGACCCCTACGACATCT 7414

QY 1242 CTGGGACTGTGTACGCGCGCTCACCTGGCCATCGATCGACGCGCGCGGAGACCAAGCGCT 1301

Db 7415 GATCGACGCGCGGCTTACTACGACAACGCGCTGGAAGCGGTCCAGATCAAGATGCCCA 7474

QY 1302 CGCCATCGGTGATGAACAATGGAAGTACGCGGACCGGAGTCCCGGACCGACATC 1361

Db 7475 CGACGTACCCCTGCAGACGCGGAGATCTACGCGCAACGCGCTTACGGGTGCGCGTCTA 7534

QY 1362 CGACGCTACAGCGGCGAAGTCTGATCTCGCGGAGTGGCGCTCGCGCGGACAGCTG 1421

Db 7535 CGCGCGCGGAGTGTGAGATCTCTGACAACTATACACGACATTCGACGACCGGTTTC 7594

QY 1422 CGGCTGTACACCTCTCGACGACCCCGTGGCAGCGCTGAGCAACTAGCTACCGCCAC 1481

Db 7595 CTACGCGGAATCCTCTCTCAGTCTCTACGACGATACCGCGCGGGGTGTCGCGCAATTCTTA 7654

|    |      |  |      |
|----|------|--|------|
| Qy | 1482 | CACCACATCCCGCAGCGAACCGTGCAGCGGACGCCCGTCTCGCATGTGAAGCAACGACGCG  | 1541 |
| Db | 7655 | CAACCAACCGGGCACCTGGATCGAAGGCAACACCACTCGTCGGCTCGGCCCAACTCCACCTA | 7714 |
| Qy | 1542 | ATACGAGATCGAGCTCGACATCGCTTGGGACACCGCGACAAAGCTTCGGCATCTCGGTGG   | 1601 |
| Db | 7715 | TGGCATCCAGGAGCGGACGAGCGGACCGACTPACAGAGCCTCTACGCCCAACAGCGTCAG   | 7774 |
| Qy | 1602 | CCGCTCCCGCAGCGAACCGGCACACGAACATCGCGCAAGTACGGAGCAGACCTGTACGT    | 1661 |
| Db | 7775 | CAATGTGCAAGACGGCTCGGTGCGCTCTACGGCGCCAACCTCCGTCTCTCCGACCTGCC    | 7834 |
| Qy | 1662 | CGACCGAGGACCTCCGACCTCGCGGGTACTCGCTCGCCGCCCTACTCGCGAGCGCGCGC    | 1721 |
| Db | 7835 | CGGCAACCGGCGACGAGCGGACACCTCGAAGGCACGGCCCGGCACACACGCTTGGCGGCAG  | 7894 |
| Qy | 1722 | CCGCATCGACCCCGCGCCCGATCGCTGCACCTGGCATCTCGTCGCACACCCAGAGCGT     | 1781 |
| Db | 7895 | CGACGCCACGAGACGCTCTCGGGCTGGACGGCAACACACCGCTTGACGCGCGCGCGG      | 7954 |
| Qy | 1782 | CGAGGTCTTGGTCAACCGCGGCACACCGTGTCTTCCAGCAGGTCCACTTTCGCCGAGGG    | 1841 |
| Db | 7955 | CAACGACATCTCGACGGCGCGCGCGGCACAACTGACCGCGCGCGCGCGCGCGCA         | 8014 |
| Qy | 1842 | CGACACGGGAATCTCGCTCTACACCGACGGCGCGCCCGCCGACACTTCACCGGCATCTCGT  | 1901 |
| Db | 8015 | CTGTTCCTCGGCTCTCGCGCGCACCGACAGCTTACCGCACCGACAGCGCCAGCTTCAACGA  | 8074 |
| Qy | 1902 | CC 1903  |      |
| Db | 8075 | CC 8076  |      |

RESULT 15  
 US-09-105-537-1  
 ; Sequence 1, Application US/09105537A  
 ; Patent No. 6265202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/105.537A  
 ; CURRENT FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 15872  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-105-537-1

|    |      |  |      |
|----|------|--|------|
| Qy | 732  | CGGGCGGTCTCGCGCTCGGACACGACCGGAGCGGCTCCGCAAGTACAGAGCA             | 731  |
| Db | 712  | CAACGATTCTCGGGGGGAGGGGGGAGTCTGCTCTACTCAAGCGCGTCTCCGCGC           | 771  |
| Qy | 792  | GTACTCTACTGTCTGACACGAGCGGGGTTCAGTTTACCGCCCTGCCACCCCGTCAT         | 851  |
| Db | 772  | CCTCGCGGAGCGGACCGTGTCCAGCGGTCTATCGCGCCACAGCGCGTCAACAACAGCAGG     | 831  |
| Qy | 852  | CGTCAACACCGAGCGGTGCGCGCGCCACCAACGCGCCGCC---GAGATCTGAGAAGCGCGAGTG | 908  |
| Db | 832  | AGCACCCCGGGTCTCACCTGCCAGCAGGCGCCGCCAGGAGAGTGTCTCGCGAGGC          | 891  |
| Qy | 909  | GTTCCGGGACCCCAAGATCCATCTGGGACACCGCCCGCGGAGATGGTGTCTGCTATCGG      | 968  |
| Db | 892  | GTACCGGAGCGCGCCTGGACCGTCCGCGCTCCAGTACGTCTGAACTTCCACGCGACCGG      | 951  |
| Qy | 969  | ACGACTCGGTTACGCGCGTTCATACCTTCGCGGAACCTGCGGACTGAGACATTCCGCG       | 1028 |
| Db | 952  | AACCCCGTCTGGGACCGCATGAGCGCCGCGCTCGCGCGCTCTTGGGCTCGGCGCG          | 1011 |
| Qy | 1029 | CAACTTCGATTACCCGAACAGCCCTCGCGCGCATCTGAGTCCCGGACCTGTTCAGAGAT      | 1088 |
| Db | 1012 | CCCGCGGAGACCCCTGCTGCTCGGCTGCGCAAGACGATCTGCGGCACTCTGAAGG          | 1071 |
| Qy | 1089 | CACCGCA---AGGAGGGACAGCGCACTGGGTGTCTCGCGCGCAGCATGAGAGCTACGG       | 1145 |
| Db | 1072 | CGCGCGCGCATCTGGGCTCATCAAGACGCTCTCTCGCTCGCGCGCGCGCGATCCC          | 1131 |
| Qy | 1146 | CATCGGCTCTCCCATGACTAGCCTACTTGGACAGGCACCTGGGACGGGAGAGATTCCA       | 1205 |
| Db | 1132 | GGGAGCGCTCAACTTCCTGACGCCCCACCGGACATCCGCTCTGACACCTCTCGGCGTCGA     | 1191 |
| Qy | 1206 | CGCGAGCACTTCACCCGCAATGGCTGCACTGGGGCTGGGACTGGTACGGCGCGCTCAC       | 1265 |
| Db | 1192 | CGTGCCGAGCGCTCTGGGAGTGGCGGACCCGCGACCGGAACTCTCTCGCGCGCTCAG        | 1251 |
| Qy | 1266 | CTGGCCATCATCAGCGCGCGGAGACCAAGCGCTCGCCATCGCTGGATGGAACAACCTG       | 1325 |
| Db | 1252 | CTGTTGGCATGGCGGCACCAAGCCCAAGTCTCTCTCAGCGAAGGCCCGCGCCAGGG         | 1311 |
| Qy | 1326 | GAGTTAGCGCGCACGGAGCTCCCAACGACGCACTCCGAGCGGTACAAACGGGCAAGACTC     | 1385 |
| Db | 1312 | CGCGAGCAGCCCGGATCGATGAGGAGACCCCGCTCGACAGCGGGCGCACTTCGCGCTT       | 1371 |
| Qy | 1386 | GATCTCGCGAGCTCGGCTCGCCGACAGCCTCGCGGCTGGTACACCTCTCTGAGCAC         | 1445 |
| Db | 1372 | CGTGTTCACCGCGCGCGGAGCGGCGCTGC---GCGCCAGGCGCGGCGCTGACAGA          | 1428 |
| Qy | 1446 | CCCGGTGGCAGCGCTCAGCAACTAGTCTACCGCCACCACTCCCGCAGCGGACCGT          | 1505 |
| Db | 1429 | GGCGTCTGAGCGGACCGGAGCTCGCGCCCGCGCACTCGCCGGTCTGCTGCACAC           | 1488 |
| Qy | 1506 | CGAGGAGCGCGCTCTGCCATGGAACGGAGCGCGCATACAGATCTGAGCTTCGACATCGC      | 1565 |
| Db | 1489 | CCGTACGGTCTTCACGACCGGTCTGTCTCTCGCCCGGACGCGCGCGCTCTCTCGA          | 1548 |
| Qy | 1566 | CTGGGACACCGGACCAAGTCTGGCATCTCGTGGCGCGCTCCCGCGACGAACCCGGCA        | 1625 |
| Db | 1549 | CGGCTCTGGGCGCTCTCGCGCGCGGACGCGCGCGCGCGGTGTCTACCGGACACCCCGC       | 1608 |
| Qy | 1626 | CAGCAACTCGGCAAGTACGGAGCAGACTGTGTACCTGTACACGAGGACCTCCGACCTCGC     | 1685 |
| Db | 1609 | CCCGGGCGCTCGCGCTCTGTTACGCGGCGAGGTTGCCCAAGCTAGGGCATGGGCAT         | 1668 |
| Qy | 1686 | CGGGTACTCGCTCGCCCTTACTCGCGAGCGCGCGCCCTCATCGACCCCGCGCGCGATC       | 1745 |
| Db | 1669 | GGAGTTGTAGCGCGCCACCCCGCTTCTCGGAGCGCTTCTGACGCGTCTCGCGCGGAAC       | 1728 |
| Qy | 1746 | GTGCACTCTGGCATCTCTGTCTGACACCAACGAGCGTCTGAGTCTTCTGTAACAGCGCGGCA   | 1805 |
| Db | 1729 | GGACCCCTCTCTGACGCGCGCTCTGCGCACTCTCGCGGCGGCGAGCACTCTCTGACCG       | 1788 |

QY 1806 CACCGTCTCTCCAGCAGGTCCACTTCGCCGAGGCGACACGGGAATCTCGCTC----- 1860  
 Db 1789 CACCGTCCACACACAGCCCGCGCTCTTCGCCGTGGAGGTCCGCCCTCCACCGCTCGTCGA 1848  
 QY 1861 -TACACCGACGCGCGCCCGCACACTTACCGGCATCGTCCGCGAGATTGGCCAGGC 1919  
 Db 1849 GTCTGGGGGCTCACGCCCGACCTGCTCGCCGCCACTCCGTGGCGAGATCAGGCCGC 1908  
 QY 1920 GATCTAGGCGATGCACACACACCGCTACCGAAGCGCGCCCGGGAGACGACGCGCGA 1979  
 Db 1909 CCACGTCCGCGGGTCTCTGCTGCGGACGCGCCCGCTCGTCCGCGCGCGCGCGCG 1968  
 QY 1980 CAATCGACAGTCTCTGTCG 1999  
 Db 1969 CCTCATGACAGGCGTCCCG 1988

Search completed: May 11, 2003, 01:58:54  
 Job time : 28690 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2003, 08:41:22 ; Search time 243 Seconds  
(without alignments)  
10233.138 Million cell updates/sec

Title: US-09-868-328B-3

Perfect score: 2001

Sequence: 1 gcggtgaccccgacttccc.....atcgacacgtctctcgctgtt 2001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description       |
|------------|-------|---------------|--------|----|-------------------|
| 1          | 97.6  | 4.9           | 2712   | 10 | US-09-748-033-4   |
| 2          | 88.2  | 4.4           | 882    | 10 | US-09-974-300-735 |
| 3          | 80.4  | 4.0           | 15872  | 9  | US-09-860-846-1   |
| 4          | 80.4  | 4.0           | 15872  | 9  | US-09-988-384B-1  |
| 5          | 80.4  | 4.0           | 15872  | 10 | US-09-861-289-1   |
| 6          | 80    | 4.0           | 1896   | 9  | US-10-124-880-15  |
| 7          | 79.2  | 4.0           | 2334   | 10 | US-09-476-242-7   |
| 8          | 78.2  | 3.9           | 11220  | 9  | US-09-860-846-32  |
| 9          | 78.2  | 3.9           | 11220  | 9  | US-09-988-384B-32 |
| 10         | 78.2  | 3.9           | 11220  | 10 | US-09-861-289-32  |
| 11         | 78.2  | 3.9           | 36778  | 9  | US-09-860-846-5   |
| 12         | 78.2  | 3.9           | 36778  | 10 | US-09-861-289-5   |
| 13         | 77.6  | 3.9           | 37948  | 9  | US-09-988-384B-5  |
| c 14       | 77.6  | 3.9           | 12441  | 9  | US-09-988-384B-3  |
| c 15       | 77.6  | 3.9           | 13613  | 9  | US-09-860-846-3   |
| c 16       | 77.6  | 3.9           | 13613  | 10 | US-09-861-289-3   |
| c 17       | 75.8  | 3.8           | 2322   | 10 | US-09-476-242-18  |
| 18         | 75.8  | 3.8           | 2322   | 10 | US-09-476-242-19  |
| 19         | 75.6  | 3.8           | 4689   | 9  | US-09-860-846-34  |

|      |      |     |       |    |                   |                    |
|------|------|-----|-------|----|-------------------|--------------------|
| 20   | 75.6 | 3.8 | 4689  | 9  | US-09-988-384B-34 | Sequence 34, Appl  |
| 21   | 75.6 | 3.8 | 4689  | 10 | US-09-861-289-34  | Sequence 34, Appl  |
| 22   | 75.6 | 3.8 | 13842 | 9  | US-09-860-846-30  | Sequence 30, Appl  |
| 23   | 75.6 | 3.8 | 13842 | 9  | US-09-988-384B-30 | Sequence 30, Appl  |
| 24   | 75.6 | 3.8 | 13842 | 10 | US-09-861-289-30  | Sequence 30, Appl  |
| 25   | 75.2 | 3.8 | 88421 | 9  | US-09-976-059-1   | Sequence 1, Appl   |
| 26   | 74.2 | 3.7 | 2322  | 10 | US-09-476-242-20  | Sequence 20, Appl  |
| 27   | 73.2 | 3.7 | 1248  | 9  | US-09-860-846-7   | Sequence 7, Appl   |
| 28   | 73.2 | 3.7 | 1248  | 9  | US-09-988-384B-7  | Sequence 7, Appl   |
| 29   | 73.2 | 3.7 | 1248  | 10 | US-09-861-289-7   | Sequence 7, Appl   |
| c 30 | 72.8 | 3.6 | 4826  | 10 | US-09-772-304B-1  | Sequence 1, Appl   |
| 31   | 69.4 | 3.5 | 1294  | 10 | US-09-748-033-2   | Sequence 2, Appl   |
| 32   | 65.8 | 3.3 | 1458  | 9  | US-09-860-846-9   | Sequence 9, Appl   |
| 33   | 65.8 | 3.3 | 1458  | 9  | US-09-988-384B-9  | Sequence 9, Appl   |
| 34   | 65.8 | 3.3 | 1458  | 10 | US-09-861-289-9   | Sequence 9, Appl   |
| 35   | 65   | 3.2 | 4257  | 9  | US-09-825-288A-1  | Sequence 1, Appl   |
| 36   | 64.6 | 3.2 | 1107  | 10 | US-09-748-033-6   | Sequence 6, Appl   |
| 37   | 64.6 | 3.2 | 1140  | 9  | US-09-860-846-15  | Sequence 15, Appl  |
| 38   | 64.6 | 3.2 | 1140  | 9  | US-09-988-384B-15 | Sequence 15, Appl  |
| 39   | 64.6 | 3.2 | 1140  | 10 | US-09-861-289-15  | Sequence 15, Appl  |
| 40   | 64.4 | 3.2 | 1266  | 9  | US-10-145-415-104 | Sequence 104, Appl |
| 41   | 64.4 | 3.2 | 3468  | 9  | US-09-988-462-2   | Sequence 2, Appl   |
| 42   | 64   | 3.2 | 2328  | 10 | US-09-476-242-6   | Sequence 6, Appl   |
| 43   | 63.2 | 3.2 | 2010  | 12 | US-10-032-717-9   | Sequence 9, Appl   |
| 44   | 63.2 | 3.2 | 2322  | 10 | US-09-476-242-5   | Sequence 5, Appl   |
| 45   | 62.6 | 3.1 | 390   | 10 | US-09-790-399-7   | Sequence 7, Appl   |

ALIGNMENTS

RESULT 1

US-09-748-033-4  
; Sequence 4, Application US/09748033  
; Patent No. US20020069431A1  
; GENERAL INFORMATION:  
; APPLICANT: Broadway, Roxanne M.  
; APPLICANT: Gongora, Carmenza E.  
; TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOIDASE AND THEIR  
; TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT  
; FILE REFERENCE: 19603/3091  
; CURRENT APPLICATION NUMBER: US/09/748,033  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2712  
; TYPE: DNA  
; ORGANISM: Streptomyces albidoflavus  
US-09-748-033-4

|                       |                 |  |            |              |
|-----------------------|-----------------|--|------------|--------------|
| Query Match           | 4.9%            | Score 97.6;  | DB 10;     | Length 2712; |
| Best Local Similarity | 43.9%           | Pred. No. 1.6e-13;                                       |            |              |
| Matches 650;          | Conservative 0; | Mismatches 794;  | Indels 36; | Gaps 4;      |
| Qy                    | 466             | CGGGCTCGTCCGTCGTCCTACACATGACGCCGCCAGCGGTCGTCTGCGACCCC    | 525        |              |
| Db                    | 845             | CGGGCTCGTCCGTCGTCCTCGTGGCTCCGTCCTCGGGGATGTCGCTC          | 904        |              |
| Qy                    | 526             | CAACGCCCGGTACACACCCAGCGGCTTACAGCTGTACTACCTGCTACCGACAGAAC | 585        |              |
| Db                    | 905             | CGCGCCCCACCCAGCGCGCGAGCGGCCGCTCTACACAGGACG               | 964        |              |
| Qy                    | 586             | AACGCCCGCGGTCGGAGCAGCAGCAGCGGTCGTCCTTACGACACAC           | 645        |              |
| Db                    | 965             | CAGGACTGGGCGCGGTCGAGGCAAGTGGAGAACACCGGACCCGCC            | 1024       |              |
| Qy                    | 646             | GGCACCGTATGCGCTCGCGCCGACTTCCCGGTGTCGGGTGTCGCGCTCGCTCGC   | 705        |              |
| Db                    | 1025            | CTCAGCGCTGAGCTTGGAGTGGGACTTCCCGCGGAGAACAGGTGACCTCGCTCG   | 1084       |              |

QY 706 ACCGGGAACACGGCAGGTTTCGGCGCCGGCGGTGCTGCGCGCTCGCGACACCGCCGACC 765  
Db 1085 GACGGCGAGCTCACCAACACGGCGACCACTGGACCGCAGAAAGAGCTGGCGGG 1144  
QY 766 GACGGGTCCGCAAGTACACGAGCAGTACCTCTACTGTCACCGACGGCGGTTTCAG 825  
Db 1145 AGCCTTGCCCGCGCGGCTCGGTGACCTTCGGCTTCAACGGCACCGGCCCGCCGACCC 1204  
QY 826 TTACCGCGCTCGCGACCGCTCATCGTCAACACCGAGGTCGGCGCCACCAACCGCC 885  
Db 1205 TCGGCTGCAAGCTCAACGGCGCTCTCGGAGCGGGGAGCTCCCGCGACACCGCG 1264  
QY 886 GCCGAGATCGAAGCGCGGAGTGGTTCCGCGACCCCAAGATCCACTGGGACACCGCCG 945  
Db 1265 CCCACCGCGCGCGACCGCCACCGCAGTACCTTCAACGAGCACTCGGTGAAGTCTCC 1324  
QY 946 GAGAGATGGTCTGGTGCATCGGACGACCTCGGTACGCGCGCTTCTACACCTCGCGA 1005  
Db 1325 TGAAGCGCGCGACGACGACGAAGGGCTCAAGAACTACGAGTCTTGGCGAGCGCGCC 1384  
QY 1006 CTGCGGACTGGACACTTTCGGCGCACTTCGACTACCCGAACAGCCCTCGCGCGC --- 1062  
Db 1385 AAGTGCACCGCTACCGCGCACCTTTCACCGACCAAGCTTCGCGCGCGCACCGAC 1444  
QY 1063 -----ATGAGTGGCGCGACCTTTCGAGTACCGCGACCGCGCGCGCGCGCG 1110  
Db 1445 TACTCTACTCGGTCCAGCGCGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1504  
QY 1111 CACTGGGTCTCGCGCGACGATGAC---GCTACGGCATCGGCTCCCATGACGTAC 1167  
Db 1505 GTCAGGTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGATC 1564  
QY 1168 GCTACTTGACAGGACCTGGGACGCGGAGAGTTCACCGCGGAGGACCTCACCCGCAA 1227  
Db 1565 AACCTCGGCTACTTACCGAGTGGGCGGTCTACGGCGCACTACCACTCAAAACCTG 1624  
QY 1228 TGGCTGAGCTGGGCTGGGACTGGTACGGCGCGGTACCTGCGGCTATCGATCGACGCGCC 1287  
Db 1625 GTGACTCGGCTCGCGCGGAGAGATCACCCACATCACTCTCTCGCAACTCCAG 1684  
QY 1288 GAGACCAAGCGCTCGCCATCGCTGGATGAACACTGGAAGTACGCGCGACGCGCTC 1347  
Db 1685 GCGGCAAGTGACCATCGGTGACAGCTTCGCGCGCTAGGCAAGCGGTACACCGCGCC 1744  
QY 1348 CCCACCGAGCTCGACGGCTTACACGGGCAAGTTCGATCGTCCGCGAGCTGGGGTCT 1407  
Db 1745 GAGTGGTTCGCGCGGTCCGCGACACTTGGGACGCGCTCGCGCGCACTTCAACCA 1804  
QY 1408 GCCGACAGCTGGGCTGGTGTACACCTCTCTGAGACCGCGCTGGCAGCGCTGACGAC 1467  
Db 1805 CTCCGCAAGCTCAAGGCCAAGTACCGGCACATCAAGGTCTCTGCTTTCGCGCGCTG 1864  
QY 1468 TACGTCACCGCGCACCAACACTCCCGACCGGACCGTTCGACGGGAGCGCC---GTCC 1525  
Db 1865 ACTGCTCGCGCGCTTACCGACCGCGGTGAAGAACCGCGCGCTTCCGCAAGTCTTCG 1924  
QY 1526 CATGGAACGAGCGCGCATACGAGT-----CGAGCTCGACATCGCCTGG 1569  
Db 1925 CACGACCTGCTCGAGACCGCGCTGGCGCGAGCTTTCGACGGCATCGACCTCGACTGG 1984  
QY 1570 GACACGGGACAGCTTGGGCTATCGGTGGCGCGCTCCCGAGCGGAACCGCGGACAGG 1629  
Db 1985 GAGTACCGGAGCGCTGGCGCTAGCTGCGACAGCTCGGTCGCGCGCGCTGAAGAC 2044  
QY 1630 AACATCGGCAAGTACGGAGCAGACCTGTACGTGACCGAGGACCTTCGACCTCGCGGG 1689  
Db 2045 ATGTTCCAGCGATGCGCGCGCGCTTTCGCGACCGGAGCTTTCGCGCGCGATCACCG 2104  
QY 1690 TACTCGCTCGCGCTTACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1749  
Db 2105 GACGCGAGCTCGCGCGGAGCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 2164  
QY 1750 CACTTGGCATCTCTGTCGACACCGCGAGCGTCTGAGGTCTTCTCAACGCGCGCGCAC 1809

Db 2165 GACTGTACAACTGATGACGTACGACTTCTTCGGCGCTGGGACAAGACCGCGCGACC 2224  
QY 1810 GTGCTCTCCACAGGTCCTCCACTTCGCGGAGGCGACACGGGAATCTCGCTCTACACCG 1869  
Db 2225 GCGCGCGCTCGCGCGCTGAACTCTTACAGCGGATCCCGAAGCGGACTTCCACTCGGC 2284  
QY 1870 GCGCGCGCGCGACACTTACCGCGCATCGCTCGTCCGCGAGA 1909  
Db 2285 GCGCGCATCGCCCAAGCTCAAGGCGGCGTCCCGCGCA 2324  
RESULT 2  
US-09-974-300-735  
; Sequence 735, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 735  
; LENGTH: 882  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-735  
Query Match 4.4%; Score 88.2; DB 10; Length 882;  
Best Local Similarity 63.4%; Pred. No. 2.4e-11;  
Matches 135; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
QY 1055 TCGCGGCGATCGAGTCCCGCGCGCTGTCGAGATCACCGACGACGCGACCGCACT 1114  
Db 654 TCGGATCATCGAATGTCGCGATCTGTATAGATGCGGCGGATGACGGAACATATAAT 713  
QY 1115 GGGTGTCTCGCCGCCAGCATGCGCTTACGGCATCGCGCTCCCGCATGACGTACGCTACT 1174  
Db 714 GGGTCTTGGAGCAAGCGGCAACGGGAAGGACCGGCAAGCGGAATCTTATGCATACT 773  
QY 1175 GGACAGGCGACTGGGACGCGGAGCAGTTCACGCGCGACGACCTCACCGCGCAATGGCTCG 1234  
Db 774 GGACAGGCGACTTCAACGGCAACGAATTTACGCGGACGACGAAGCAACCCCGTGGCTTG 833  
QY 1235 ACTGGGCTGGGACTGGTACGCGCGCTCACCT 1267  
Db 834 ATTACGGATTGACTGGTATGCGGACGTGACAT 866  
RESULT 3  
US-09-860-846-1  
; Sequence 1, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860,846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match      4.0%; Score 80.4; DB 9; Length 15872;
Best Local Similarity 42.7%; Pred. No. 1e-09;
Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCCCCAGCGGTGCTCTGACACCCCAAGCCCGGTCACACACCGGCGC 551
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 472 CATGCCCAACCGCGTCTGTTACACCTCGGCTGAGGGCCGAGCCTCACCGTCGACG 531
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 552 CTACACGCTGTACTACCTGCACTGACAGCAACAGCCGCCCGGGTGGAGCACCGC 611
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 532 CGCGAGTGTCTCGCTGTCGCGTGCAGCTGCGGCTGCGAGTCTCGCGCGCGGGGA 591
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 612 GAGCAGACCGGCGGCTGCTTACAGCACACAGCGACCGGTCGATGCGGCTCGGCGCG 671
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 592 GTCCACGACGGCGCTGTCGCGCGGCTGAACCTCAACATCTCTCGCGGAGCGCGTGAC 651
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 672 CTTCCCGTGTGTCGCGGTGCGGCTGTCGGACCGGGAACAGCGAGGTTTCGGGCG 731
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 652 GGAGGAGGCTTTCGGTGGACTCTCCCGGACGGCACCGCTTACGCTTCGAGCGCGGGC 711
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 732 CGGCGCGTGTGTCGCGCTGCGACACAGCCGACCGCGCTCCGCAAGTACGAGGACA 791
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 712 CACGGATTTGTCGCGGGGAGGGCGGAGTGTCTGCTACTCAAGCGGCTCTCCGCGCG 771
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 792 GTACCTTACTGTGTCGACGCGGCGGTTTACGTTTACCGGCTTCCCGACCGCGTCTAT 851
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 772 CCTCGCGGACGCGACCGTGTCCACGGCTCATCCGCGCGCGCGTCAACACGACGG 831
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 852 CGTAAACACGAGTGTGCGCGCGCCACACGCGCGCGCGCGCGCGCGCGCGCGCG 908
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 832 AGCCACCGCGGCTTTCACGCTGCCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCG 891
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 909 GTTCCGCGACCCCAAGATCCACTGGACACCGCGCGCGGAGATGGTCTGCTGATCGG 968
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 892 GTACCGGAGGCGCGCTTGGACCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCG 951
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 969 AGGATGCGGTACGCGCGGTTTACACCTGCGCGGACCTGCGCGACTGAGACACTTCGCG 1028
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 952 AACCCCGCTGCGCGACCCATCAGAGCGCGCGCGCTGCGCGCGCTTCTCGGCTGCGCG 1011
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1029 CAACCTTCGACTACCGACCGACCGCTTGGCGGATCGAGTCCCGGACCTGTTGAGAT 1088
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1012 CCGCGCGGACGAAACCCCTGCTGCTGCGGTGCGGCAAGACGAACTGCGGCGACCTCGAAG 1071
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1089 CACCGCAG---ACGACGGGACACGCACTGGGTGCTGCGCGGACGATGAGCGCTACGG 1145
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1072 CGCGCGCGCATCTGCGGCTCATACAGCGTCTTCTCGGCTGCGCGCGCGGATCCG 1131
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1146 CATCGGCTCTCCCATGATGCTTACGCTTGGAGAGGACCTTGGAGCGGCGAGCACTTCA 1205
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1132 GCGGAGCTTCACTTCCGTACGCGCCACCGGACATCCGCTCGACACCTCTGGGCTCGA 1191
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1206 CGCGGACGACTCACCGCGAATGCTGATGCGGCTGAGTGTAGCGGCGCGCTAC 1265
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1192 CGTGGCGGACGCGCTTGGGAGTGGCGGACCGCGGACCGGAACTCTCTCGCGGCGTCA 1251
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1266 CTGGCATCGATCGAGCGCGGAGACCAAGCGCTGCGCATCGCTGATGAACAATG 1325
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1252 CTCGTTGGCATGCGGCGGACCAACGCGCTGCTGCTCTCAGGAAGGCGCGCGCGAGG 1311
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1326 GAAGTACCGCGACGAGCTGCGCGGACGACGATCCGAGCGCTACACGGGCGAGAACTC 1385
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1312 CGCGGAGGACCGCGCATGATGAGGAGACCCCGCTCGACAGCGGCGCGCACTGCCCT 1371
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1386 GATCTCGCGAGTGTGCGGCTGCGCGGACAGCTTGGCGGTGATACACCTCTCTGAGCAC 1445
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 4

```
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1
```

```
Query Match      4.0%; Score 80.4; DB 9; Length 15872;
Best Local Similarity 42.7%; Pred. No. 1e-09;
Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCCCCAGCGGTGCTCTGACACCCCAAGCCCGGTCACACACCGGCGC 551
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 472 CATGCCAACCGGCTCTCGTACCACCTCGCGCTGCGAGGCCCGGAGCCTCACCGTCCAGCG 531  
QY 552 CTACAGCTGTACTACTCTGACTCCAGCAGAACAGGCCCCGGGTGGGACACAGC 611  
Db 532 CGCGCAGTGTCTCTGCTGCTGCGCTGACCTGAGTCCCTGCGCGCCGGGA 591  
QY 612 GAGCAGCAGCGGCGTCTGCTTACGACACACGACGCTGATCCGCTGGGCCGA 671  
Db 592 GTCCACGAGCGGCGCTCTGCGCGGCTGAACTCAACATCTCGCGGAGAGCGCGTGAC 651  
QY 672 CTTCGCCGTGTGTGCTGGCGTCTGCGGACCGCGAACACCGGAGGTTTCGGCG 731  
Db 652 GGAGGAGCGCTTCGCTGAGTCTCCCGGAGGACCGCTACACCTTCGACGCGCGG 711  
QY 732 CGCGCGCGTCTGCTGCGCTCGCGACCGACGACCGCGGCTCGCGAGTACAGGAG 791  
Db 712 CAACGATTCGTCGCGGCGGAGGCGGCGGAGTCTGCTACTCAAGCGCTCTCCGCG 771  
QY 792 GTACCTTACTTGTTCAGCAGCGGCGGTTTCAGTTTCAGCGGCTCGCGGACCGCTCAT 851  
Db 772 CCTGCGCAGCGGACCGCTGTCTCCAGCGGCTCATCGCGCTACCGCGGCGCTCAACAGCAGCG 831  
QY 852 CGTCAACACCGAGTCTGCGCGCGCGCACACCGCGCGC---GAGATCGAGAACCGGAGTG 908  
Db 832 AGCCACCGCGGCTTACCGTGCCTCCAGCAGCGCGCGCGAGGAGTCTGCGGAGCG 891  
QY 909 GTTCGCGACCGCCAGATCACTGGGACACCGCGCGGAGATGGGTCTGCTCATCGG 968  
Db 892 GTACCGGAAGCGGCGCTGAGCGCGTCCGCGCTGCTGAGTACGTAACACCGCACCGG 951  
QY 969 AGGACTGCGGTAGCGCGGCTTCTACACTCGCGGAGAACCTCGCGGACTGGACACTTCGCGC 1028  
Db 952 AACCCTGCTGCGGACCGCATGAGCGCGCGCGCTCGCGCGCGCTCTCGGCTCGCGCG 1011  
QY 1029 CAATTCGACTACCGGAACCGCCTCGCGCGCTGAGTGCCTCGCGACTCTGTTCGAGAT 1088  
Db 1012 CCGCGGAGAGACCGCTCTGCTGCTGCGCGCGCGAGAGAGTCTGCGGACCTCGAAG 1071  
QY 1089 CACCGCAG---ACGAGGAGACCGCCTACTGGTGTCTGCGCGGACGATGAGCGCTACGG 1145  
Db 1072 CGCGCGCGCATCGTGGCGCTCATCAAGCGCTCTCGCGCTCGCGCGCGCGGATCC 1131  
QY 1146 CATCGGCTCCCATACGCTACGCTTACTGGACGAGCCTGGGAGCGGCGGACAGTTCA 1205  
Db 1132 GCGGAGCTCAACTTCGTACGCGCGCACCGGACATCCCGCTCGACACCTCGGCGTCA 1191  
QY 1206 CGCGGACGACCTCACCGCCCAATGCTCGACTGGGCTGGGCTGCTAGCGCGCGCTCAC 1265  
Db 1192 CGTCCCGAGCGCTCGGAGTGGCGGACCGCGACCGGACCGGACTCTCGCGCGGTCA 1251  
QY 1266 GTGGCGCATCGATCGAGCGCGCGAGACCAAGCGCTCGCGCATCGGCTGGATGAACAATG 1325  
Db 1252 CTCGTTTCGCGATGGCGGCGGACCAAGCGCGCACCTGCTCTCGAGGAGGCGCGCGG 1311  
QY 1326 GAAGTACCGCGGACGAGTCCCGACCGACCGCATCCGAGCGCTACAGCGGACAACTC 1385  
Db 1312 CGGCGAGACGCGCGCATCGATGAGGAGACCGCGCTCGAGCGGCGCGGCGCTCCCTT 1371  
QY 1386 GATCGTCCGAGCTGCGGCTCGCGGACGAGCTGCGGCTGGTACACCTCTCTGAGCAC 1445  
Db 1372 CCGTTCACCGCGCGGCGGAGGCGCTG---GCGCGGAGCGCGGCGCTCGACGA 1428  
QY 1446 CCGCGTGGAGCGGTGAGAACTAGTACACCGCGGACCGCATACGAGTCTGAGCTCGCATCGC 1505  
Db 1429 GCGCGTCAAGCGGAGCGGAGCTCGCGCGCGCGGCTCGCGCGCTGCTGGTCAACAC 1488  
QY 1506 CGAGGCGAGCGGCTCTGCTCCATGAAAGGAGCGCGCATACGAGTCTGAGCTCGCATCGC 1565  
Db 1489 CCGTAGGCTTTACGACACCGGTCTGCTGCTCGCGCGGACCGCGCGCGCTCTCGA 1548  
QY 1566 GTGGGACACCGGAGGAGTCTGCGATCTCGGCTGGCGCGCTCCCGCGGAGGAGGAGGAGTTCGGCGC 1625  
Db 1549 CGGCGCTCGCGCGCTCGCGCGGAGCGCGCGCGCGGCTGCTACCGGACCGCGCGC 1608

## RESULT 5

US-09-861-289-1

; Sequence 1, Application US/09861289

; Patent No. US200110897A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; FILE REFERENCE: 600.438U1

; CURRENT APPLICATION NUMBER: US/09/861,289

; CURRENT FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 15872

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-861-289-1

## Query Match

; Score 80.4; DB 10; Length 15872;

; Best Local Similarity 42.7%; Pred. No. 1e-09;

; Matches 649; Conservative 0; Mismatches 856; Indels 15; Gaps 4;

QY 492 CATGACGCCGCCAGCGGCTGCTGCGACCGCCCAACCGCGGTCACACCGCCAGCGCGC 551

Db 472 CATGCCAACCGGCTCTGTAACCTCGGCTGCGAGGCGCGGCTCACCGTCAACGTCAGCGC 531

QY 552 CTACAGCTGTACTACTCTGCACTCGGACGACGAGAACGAGCGCGCGGCGGTGGGACACGC 611

Db 532 CGCGCAGTGTCTGCTGCTGCTGCGCTGCACTGCGGCTGCGAGTCCCTCGCGCGCGGGA 591

QY 612 GAGCAGCAGCGGCGGCTCTGCGACCGACCGACCGCGGCTGATCGCGCTCGCGCGCGCA 671

Db 592 GTCCACGAGCGGCGCTGCTGCGCGGCTGAACTCAACATCTCGCGGAGAGCGCGCTGAC 651

QY 672 CTTCGCCGTGTGCTCGGCTCGCGGCTGCTGCGGACCGCGCGCGCTTACACTTCGACGCGCGGCGC 731

Db 652 GGAGGAGCGCTTCGCTGGAGACTCTCCCGGACGCGCGCGCTTACACTTCGACGCGCGGCGC 711



```
QY 732 CGGCGGGTCTGTCGGCTCGCGGACCCAGCCGACCGAGCGGGTCCGCAAGTACCGAGGACA 791
Db 712 CAACGGATTCTCGCGGGGCGAGGGCGGAGTCTGCTTCTGCTACTCAAGCGGCTCTCCCGGCG 771
QY 792 GTACTCTACTGTTGTCGACGCGGGGTTACGTTTACCGCTTCCCGGACCCCGTCTAT 851
Db 772 CTTGCCGACGGGACCGTGTCTACGGGCTATCCGCGCCAGCGGCTTCAACAACGACGG 831
QY 852 CGTCAACACCGACGGTTCGCGGCCACCAACGCCCCGCC---GAGATCGAAGACCGCGAGTG 908
Db 832 AGCCACCCGGGTCTCACCGTGCACAGCAGGAGCGGCCCCAGAGAGGTGCTTCGCGGAGCG 891
QY 909 GTTTCGCGACCCCAAGATTCACCTGGGACACCGCCCGCGGAGAAATGGTCTGCTCATCGG 968
Db 892 GTACCGGAAGCGGGCCCTGGACCCCTCGCGGCTCCAGTACGTCGAATCCACGGCACCGG 951
QY 969 ACGACTGCGGTACGCGCGTTCACACCTCGCCGAACTCGCGGACTGGACACTTCGCGG 1028
Db 952 AACCCCGTTCGGGACCCATCGAGGCGCCGCGCTCGGCGCGTCTTCGGGTCGGCGCG 1011
QY 1029 CAACTTTCGACTACCCCGAACCAGCCCTCGGGGCACTGAGTGCCTCCGACCTGTTTGGAGAT 1088
Db 1012 CCCGCGGACGACCCCTGCTCTGGGCTCGGCCAAGACGAGCTCGGCGACCTCGAAGG 1071
QY 1089 CACCGCAG---ACGACGGGACACGCCACTGGGTGCTCGCGCCAGCATGGACGCTACGG 1145
Db 1072 CGCGCGCGGCACTCGTCGGCCCTCATCAAGACGCTCTCGCGTTCGCGCGCGCGGATCCC 1131
QY 1146 CATCGGCTCCCATGACGTAGCTAGCTACTGACAGGCACTGGGACGGGAGCACTTCCA 1205
Db 1132 GCGAGGCTCAACTTCCGTAGCCCCACCGGACATCCGCTCGACACCTTCGGGCTCGA 1191
QY 1206 CGCGGACGACTCACCCCGCAATGCTGACTGGGGCTGGGACTGGTACGGCGCGCTCAC 1265
Db 1192 CGTGCACGCGCTCGCGGAGTGGCGCACCCGACCGCGGAACCTCTCGCGGGGCTCAG 1251
QY 1266 CTGGCCATCGATCGACGCGCGGAGACCAAGCGCTCGCCATCGCGTGGATGAACAAGT 1325
Db 1252 CTGCTTCGGCATGGCGCGACCAACGCGGACGCTGCTCTCAGCGAAGCGCCGCGCCAGG 1311
QY 1326 GAAGTACGCGCACGCGACGTCCCAACGACGATCCGACGGCTACAACGGGCAACATC 1385
Db 1312 CGGCGAGACCGCGCATCGATGAGGAGACCCCGCTCGACAGCGGGCGGCTACTGCCCTT 1371
QY 1386 GATCGTCCGCGAGTTCGGCTCGCCGACAGCTTCGGGCTGGTACACCTTCCTTGAGCAC 1445
Db 1372 CGTCTCACCGCGCGCGCGGAGGCGCTGCG---GCGCCAGGCGCGCGCTTCACGA 1428
QY 1446 CCCGTGCGAGGCTGACGAACTAGTCAACGCGCCACACGACACTTCCCGGACCGGACCGT 1505
Db 1429 GCGCGTCAAGCGGACCGCGGAGCTCGCGCGCGCGGCACTCGCGCGGCTGGTTCACCCAC 1488
QY 1506 CGAGGCGACCGCGTCTGCCATGGAACGCGCGGCATACAGATTCGAGCTCGACATCGC 1565
Db 1489 CCGTACGGTCTTCAGCACCGGTCGGTCTGCTCGCCCGGACCGCGCGGCTCTCTCGA 1548
QY 1566 CTGGGACACCGCGAGAGCTCGGATCTCGGTGGCGCGCTTCCCGGCAACCGGCA 1625
Db 1549 CGGCTCGCGCGCTCGCGCGGACGCGCGCGCGGCTGGTTCACGCGACCCCGCG 1608
QY 1626 CACGAACATCGGCAAGTACGAGACGAGCTGTACGTGACCGAGGACCGCTCGGACCTCGC 1685
Db 1609 CCGCGGCGCTCGCGGCTCTGTTTCAGCGGCGAGGTTGCCAAGCTACGGGCTATGGGAT 1668
QY 1686 CGGGTACTCGTCCGCCCTACTCGGAGCGCGCCGCCCATCGACCCCGGCGCGCGATC 1745
Db 1669 GGAGTTGTACGCGCGCCACCCCGCTTCGCGACGCGCTTCGACGCGCTTCGCGCGGCGA 1728
QY 1746 GTGCACTTCGCACTCTCTGTCGACACCCAGAGCGCTTCGAGGCTTCTGTCACCGCGGCA 1805
Db 1729 GGACCCCTCTCGACCGCGCGCTCGCGGAACTGTCGCGGGGCGGACACCTCGACCG 1788
QY 1806 CACCGTGTCTCTCCAGCAGGTCCACTTCGCGGAGGCGGACACGGGAATCTCGCTC---- 1860
```

```
Db 1789 CACCCTCCACACACAGCCCGGCTCTTCCGCTGGAGGTTCGCTCCACCGCTCTGTCGA 1848
QY 1861 -TACACCCAGCGCGCGCGCACACTTCACCGGCATTCGTCGCGGAGATTGGCCAGCG 1919
Db 1849 GTCTTGGGCGTACGCGCGGACCTGCTCGCGGCCACTCTCGTCGGGAGATCAGCGCGC 1908
QY 1920 GATCTAGGCGATGACACACACACCGCTCACCGAAAGCGCGCGCGGAGACGACGCGCA 1979
Db 1909 CCACGTGCGCGGGTCTTGTCTGCTGCGGAGACGCGCGGCTCTGTCGCGGCGCGGCG 1968
QY 1980 CAATCGACACGCTCTCTGTCG 1999
Db 1969 CCTCATGAGCGCTCCCG 1988

RESULT 6
US-10-124-880-15
; Sequence 15, Application US/10124880
; Publication No. US20030026810A1
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene No. US20030026810A1boe
; APPLICANT: Schulein, Martin
; APPLICANT: Outtrup, Helle
; TITLE OF INVENTION: No. US20030026810A1el Rhamnogalacturonan Hydrolases
; FILE REFERENCE: 5572.204-US
; CURRENT APPLICATION NUMBER: US/10/124,880
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/311,626B
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 0608/98
; PRIOR FILING DATE: 1998-05-01
; PRIOR APPLICATION NUMBER: 60/084,358
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-124-880-15

Query Match 4.0%; Score 80; DB 9; Length 1896;
Best Local Similarity 43.8%; Pred. No. 1.6e-09;
Matches 665; Conservative 0; Mismatches 825; Indels 30; Gaps 6;

QY 489 CCACATGAGCGCCCCAGGGCTGCTGTGCGAACCCCAACCGCCGGTCCACCCACCGG 548
Db 276 CAACTCCAGCCCGTACCGGCTCCACCACTTCTTCCACTCCGCGCGCTCCGACGC 335
QY 549 CCGCTTACCAGTGTACTACTCTGCTGCGACAGAACACGCGCCCGCGGGTGGGACCA 608
Db 336 CGACTTACACCGTCCGCGGGTCTGTAACGCGACGAGGCGGACTCGTGTCCACGCGAT 395
QY 609 CCGGAGACGACGACGCGGCTGCGCTTACGACCAACGCGACCGCTGATGCCCTGCGGC 668
Db 396 CCAGTTCGGGCGGCTACAGGAGCTACCGATCAGCCCGCTCCGGGCGACCAACCC 455
QY 669 CGACTTCCCGGTGTTGGTCCGGGTGCGGGTGTGTCGGCACCGGCAACAGCGAGGTTGCG 728
Db 456 CGACGGGCTCTCTACACCTAGAGGCGCAACGACGCTCCGTCGGGAGCTTCGAGCGGCA 515
QY 729 CCGCGGGGCTGCTGCTGCGCTCGGACCGGACCGCGGCTCCGCAAGTACACCA 788
Db 516 CGCGCCCTTCGACCTGCTTCAAGTGGCGCGCCACCAACGCAAGCACTCCAGTC 575
QY 789 GCAGTACCTCTACTGTCGACCGAGCGGGTTCAGTTTACGCTTCCGCTCCCGACCGCT 848
Db 576 CGGCTACACCGGCAACGCGTCTGACGGGATCAGCTCGACGGA---CCGCGCTGTG 632
QY 849 CATGCTCAACCGGAGGTTCGCGCGCCACCAACGCGCGCGGAGATCGAGACCGCGAGTG 908
```

Db 633 GCGGCTCGACCTGGGCGCGAACAATCGCTCCGGCGCCCACTACACCCAGTTCCAGTGTA 592  
QY 909 GTTCCCGCGACCCCAAGATCACTGGGACACCGCCCGGAGAAATGGTCTGCTCATCGG 968  
Db 693 CGACTAGGAGCGGCGCGCGCGGCGAGTGCCTATGAGACCGCGACGCAACCAAGA 752  
QY 969 AGCACTGCGGTAGCGCGCGTCTTACACCTCGCGGAACCTGCGGACTGGACACTTCGCG 1028  
Db 753 CGGCACCGCGCGGTATCGGCAACTCTCTCGCGGATCACCGCACTCGAGCGGTACGT 812  
QY 1029 CAACTTCGACTACCCGAACACCGCCTCGGGGCGATCGAGTGGCCGACCTGTTCCAGAT 1088  
Db 813 CTTCTCGCGCGCGAATACCTCACTGTTCAACGCGCGGACCGGACCGATGGGAC 872  
QY 1089 CACCGCAGACGAGGACACGACCTGCTGCTCGCC---GCCAGATGACGCTACCG 1145  
Db 873 CTGCGACTAGTCCCGCGCGCGGTCTGCTCTCTCTGGGCGACTCTACGGCAACCG 932  
QY 1146 CATCGCGCTCCCATAGCTAGCTTACCTTGGACAGGACCTGGGACCGGAGAGTTCCA 1205  
Db 933 CGTCGACCGGTTCTGGCGGCGACGCGTACCTGGACGCTCCCGCCCTCGGTGATTAT 992  
QY 1206 CCGCGACGACCTCACCCCGCAATGGCTCGACTGGGCTGGGACTGTACGCGCGCTAC 1265  
Db 993 GCGCGCGGTACTACCGCGGACGCTGATCGCGGCTGGGACTGG---CGGACCGCG 1049  
QY 1266 CTGGCCATCGATCGAGCGCGCGGAGACCAAGCGCTCGGCATCGCTGGATGAACAATG 1325  
Db 1050 GTTACCGCGCGCTGGACCTTCGACACCAACTCTCCACCAACAGCGGCAAGGCTACGA 1109  
QY 1326 GAAGTACCGCGGACGCGCTCCCAACGACGATCCGACGCTGACAGCGGAGAGTCA 1385  
Db 1110 CGCGCAGGCAACACACAGCTCTCGCTCGGACGCTGGAGCTGACGCGCGGACGAGAT 1169  
QY 1386 GATCGTCCGCGAGCTGGGCTCGCCGACAGCTGGCGCTGTACACCTCTGAGCAC 1445  
Db 1170 CGTCTACGCGCGGATGGCGCTCGAGGACAAACGCTACGCGCTGTGGACCAACAGGACGG 1229  
QY 1446 CCGCGTGGCAGCGCTGACAACTAGTCAACCGCCACAC---ACACTCCCGGACCG 1499  
Db 1230 CCACGCGACGCCATGCAGCTCGGCGACCTCGACCGTCCCGGCGGCGCTGAGGAGTT 1289  
QY 1500 GACCGTACGAGCGCGCTCTCGCATGGAACGACGCGCATAGGATGAGCTCA 1559  
Db 1290 CAACTTCGACGAGGAGCGCTGAAAGCGCTCGTACCTGGGCGACGCGCGACGGCA 1349  
QY 1560 CATCGCTGGGACACGCGGACGATCGGCATCTCGGTGGCGCTCCCGCGACGGAAC 1619  
Db 1350 GATCCTCTGTTCCACCGCGCGGCGGACAAAGCGCGGTCTCGGGGACATCTG 1409  
QY 1620 CCGGCACACGAATCGGCAAGTACGAGACAGACCTGTACGTCGACCGAGGACCTCCGA 1679  
Db 1410 GTCCGCGACGCGCGCGCGAGTCTCTGTC---GTCCGCGAGAGCGCATCCGCA 1463  
QY 1680 CTTCCCGGCTACTCGCTCGCCCTACTCGGAGCGCGCGCCCACTACACCGCGCGC 1739  
Db 1464 CCCCAGGCGACCGTCTCGCGAGCGGCAAGCGCTCCAGCGCAACTCTCTTCTGTTG 1523  
QY 1740 CGGATCCGTCACCTCGCGCATCTCTGTCACACCGACGAGCGTCTGCTGCAACGC 1799  
Db 1524 GGACGGGACACCGTCCGTGAACTCTCGACGGCACCCACGTCGACAACTACGCACTTC 1583  
QY 1800 CGGCGACACCGTCTCTCCAGAGTCCACTTCGCGGAGGCGACACGGGAATCTCGCT 1859  
Db 1584 GGGCGACA-----CCCGCTGTCTACCGGCTCCGCGCTCTCCAAACAGGCGAC 1634  
QY 1860 CTACACCGAGCGCGCGCGACACTTCACCGCATCTGCTCGCGGAGATTGGCGAGC 1919  
Db 1635 CAAAGCCACCGCGCTCTGGCGCGGACATCTCGCGGACTGGCGGAGGCTGCTG 1694  
QY 1920 GATCTAGGCGATGCACACACAGCTACCGAAGCGCGCGCGCGGAGACGACGCGCA 1979

Db 1695 GCGCAGCTGAACAACACGCGCCCTGCGATCTACTCCACCCCTACGACAGGACCCG 1754  
QY 1980 CAATCGACACGCTCTCGTGC 1999  
Db 1755 CATCAGCACCTCTCTCCAG 1774  
RESULT 7  
US-09-476-242-7  
; Sequence 7, Application US/09476242  
; Patent No. US20020146683A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: MARTIN, Karin  
; APPLICANT: HARTOG, Eric  
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
; FILE REFERENCE: 1605,002  
; CURRENT APPLICATION NUMBER: US/09/476,242  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 2334  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Leu122-Ser199  
US-09-476-242-7

Query Match 4.0%; Score 79.2; DB 10; Length 2334;  
Best Local Similarity 43.8%; Pred. No. 2.4e-093;  
Matches 458; Conservative 0; Mismatches 573; Indels 15; Gaps 2;

QY 795 CTTCTACTGCTGACGACGCGGGGTTACGTTACCGCCCTGCCGACCCCGTCATCGT 854  
Db 147 CTGTCTCGCGCAGGACGCCAAGGCCTACGACACCGAGGTGCACAACGTGTGGGCCAC 206  
QY 855 CAACACCGAGGTTCGCGCGCCACACGCGCCGAGATCGAGAACGCGGAGTGTTCG 914  
Db 207 CCACGCTCGGTGCCACCGACGCCCAACCCCGAGGAGATCGTCTGGAGAACGTGACCGA 266  
QY 915 CGACCCCAAGATCCACTGGGACACGCGCGGAGATGGTCTGCTGATCGGACGACT 974  
Db 267 GAATTTCAACATGTGAAGAACAACATGTGGAGAGATGACGAGGACATCATAGCT 326  
QY 975 GGGGTACGCGGGTCTACACCTCGCGAACCTCGCGGACTGACACTTTCGCGCAACTT 1034  
Db 327 GTGGACACGAGCTGAAGCCCTCGTGAAGCTGGGCAACAGCGTATCACCCAGGCTG 386  
QY 1035 CGACTACCGGACCGCCCTCGGGGCGATCGAGTCCCGGACCTGTTTCGAGATCACCG 1094  
Db 387 CCCCAGGTGAGCTTGGAGCCCATCCCATCTACTGCGCCCGCGCGGCTTCGCCA - 445  
QY 1095 AGACGACGGGACACGACCTGCTGCTCGCGGACGCTGAGCGCTACGGATCGGCT 1154  
Db 446 -----TCCTGAAGTGCACGACGAAGAGTTCACGCGACGCGGCGCTGACCAACGAG 500  
QY 1155 CCCCATAGCTAGCGCTACTGACAGGCACTTGGGACGCGGAGAGTTCACCGCGGACGA 1214  
Db 501 CACCGTGCAGTGCACCCACGGCATCGCGCGGTGGTGGAGCACCCAGCTGCTGTGAACGG 560  
QY 1215 CTTACCCCGCAATGGCTGACTGGGCTGGGACTGGTGGTGGGCGGCTACCTGGGCCATC 1274  
Db 561 CAGCTTGGCGGAGGCGGTGTGATCGCGACGAGAACTTACCGCAACAGCCCAAGAC 620  
QY 1275 GATGACGCGCGCGGAGAACCAAGCGCTCGCCATCGGCTGGATGAACAACCTGGAAGTACGC 1334  
Db 621 CATCATCTGTCAGCTGAAGAGAGCGTGGAGATCACTGCACCCCGCCCAACACACAC 680  
QY 1335 CGACGCGAGCTCCCGACCGGACGATCCGACGCGCTACACGCGGAGAGTTCGATCGTCCG 1394  
Db 681 CCGCAAGAGCATCACCATCGGCGCGCGCTTCTACGCCACCGCGGAGACATCATCG 740



Db 4000 GTGACCTGGTGGCCCTGGAGCCCTGGGAGGCGACGCCGGTCACCGAGGGTCGGACCGGG 4059  
QY 1780 GTGAGGCTTTCGTAACGCCGGCCACACCGCTCTCCAGCAGGTCCACATTCGCCGAG 1839  
Db 4060 GAGCGGTGCGCGCTGGCGCTGGCCCTCGCCCGCGACGCGCTCACCGCCCTG 4119  
QY 1840 GCGGACCGGGAATCTGGTCTACACGACGCGCGCCCGGACACATTCACCGG 1892  
Db 4120 GACACCGCGCTGGCCACGCGGACACCGCGCTCAGGATCGCGGAGGTGACTG 4172

RESULT 9  
US-09-988-384B-32  
; Sequence 32, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 32  
; LENGTH: 11220  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-988-384B-32

Query Match 3.9%; Score 78.2; DB 9; Length 11220;  
Best Local Similarity 41.9%; Pred. No. 3.4e-09;  
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCGTCTGCTCGGCTGTACACATGACGCCCCCAGCGGTGCTCTGGACACCCCAA 528  
Db 2740 GACTCCCGCGCGTCCAGGGCTCCGTGCAGGACTCCTGGCGTACCGACTGGAAG 2799  
QY 529 CGCCGGTCAACACCCAGCGGCGCTTACCAGTGTACTACCTGCATCTCGACGAGAAC 588  
Db 2800 CGCTCGCGGTCCGCGACGCGTCCAGCGCGCGGCTGTCCGGCGTGGCTCTGCTC 2859  
QY 589 GGCCTCGCGGTGGGACACCGAGACGACGCGGCTGCGCTTACGACACACCGC 648  
Db 2860 GTCCCGGAGGACCTTCGCGGAGCGCGCGGCTGTGCTGCGCGCTGTCCGGCGCGC 2919  
QY 649 ACCGTGATGCCGCTCGCGCCGCTTCCCGTGTGGTTCGGGTGCGGCTGTGCGGAC 708  
Db 2920 GCGACCCCGTACAGCTGAGAGTGTCCCGCTGTGGCGAGCGGACGCGCTCGCCGCG 2979  
QY 709 GCGACACGCGC---AGGTTTGGCGCGCGGCGGTGCTGCGGCTCGGACCGCGGAC 765  
Db 2980 CTGGCGGAGGCGCTGGCGGCGCGCTGGAGCGCTGACGCGGCTCTCTGCTGCTCG 3039  
QY 766 GAGCGGCTCGCACTACACGAGCAGTACCTCTACTGTCGACGCGGCGGTTCAG 825  
Db 3040 TGGGACGAGAGCGCGCACCCCGCGCACCCCGCGCTTACCCCGGGGACCGGCGCC 3099  
QY 826 TTCACCGCGCTCGCGGACCGCTCATCTCAACACCGAGGTGCGCGCGCGCACCGCC 885  
Db 3100 CTCACCTGCTGACGCGCTGAGGACGCGCGGCTGCGCGCGCTGTGGTGGTGACC 3159  
QY 886 GCGGAGATCGAGACGCGCGAGTGTTCGCGGACCGCCCAAGATCCACTGGGACACGCGCGC 945  
Db 3160 CACGCGCGGCTGTCTGCTGGCGGCGCGCGCTGACCTCTCCCGCGCGGCGGCTGTG 3219  
QY 946 GGAGAATGGTCTGGGTATCGGAGACTCGGAGTACCGCGGCTTCTACACCTCGCGGAC 1005

Db 3220 TGGGCACTGGGCGGCTCGCCGCCCTGGAGGACACCCCGAGCGGTGGGGCGGCTGTATCGAC 3279  
QY 1006 CTGCGCGACTGGACACTTTCGCGCGCAACTTCGACTACTCCGGAACACCGCCTCTCGCGCGCATC 1065  
Db 3280 CTGCGCTCGGAGCGGACCGGGCGGCTTGGACCGCATGACACAGTCTCTCGCGCGGCT 3339  
QY 1066 GAGTCCCGCGGCTTTCGAGATCACCGAGACGAGGAGACGCGCACTTGGGTGCTCGGC 1125  
Db 3340 ACGGTGAGGACAGGTGCGGTACGCGCTCGCGGCTGCTCGCGCGCGCTCTCTCGCG 3399  
QY 1126 GCCAGCTGAGCGCTACGGCATCGGCTCCCATGAGTACGCTACTTGGACAGGAC 1185  
Db 3400 GCTTCCCTCCCGGCGCACGCGACGGCTTTCGCGCTGTGGCAGGCGGAGGCGGCTGCT 3459  
QY 1186 TGGGACGCGGAGCTTTCACGCCGACGACCTTACCCCGCAATGGCTCGACTGGGCTGG 1245  
Db 3460 GTCACCGGTGCGGAGGAGCTTGGCGCGCGGAGCGCGGCTGGCGCGGACGCGC 3519  
QY 1246 GACTGTAGCGGCGCTCACTTGGCCATCGATGAGGCGCGGAGACCAAGCGCTCGGC 1305  
Db 3520 GCGGACACCTCTCTCTCCACACACCCCTCCGCGCAGGAGGCGCGGAGGCGGACCTCC 3579  
QY 1306 ATCGGTGATGAACAATGGAAGTACCGCGACGCGAGTCCCGACGACGACGACATCCGAC 1365  
Db 3580 GGTGCGCGGAGGACTTCCGCGCTCCCGGGCTGTGCGGAACTCGCGGACCTGGGCGG 3639  
QY 1366 GGCTACAAAGCGGCGAGCTCGATCGTCCGCGAGCTGCGGCTCGCCGCGACAGC-----CT 1419  
Db 3640 ACGGCGACCGTGTGACCTTCCGACCTACGACGCGGAGGCGCGGCGGCTGTCTCGC 3699  
QY 1420 GCGGCTGTACACCTCTCTGAGCACCCCGGTGGGAGGCTGACGAACTACGTCACCGC 1479  
Db 3700 GCGGCTCTCCGCGCGACCGCTACGCGCGCTCTTCCACCTGCGCGCGGCGGCTCGACTCC 3759  
QY 1480 ACCACACATCCCGGACCGGACGCTCGAGCGCGCTCTCGCTGCGGCGGCGGCTGCGG 1539  
Db 3760 GAGCGCTCGCGCGACGCGGACGCGCTGCGCGCTGTGCTGACCGGAGGCGGCGC 3819  
QY 1540 GCATACGAGATCGAGCTCGCATCGCTGGGACACCGCGAGCAAGCTCGGCTCTCGGTG 1599  
Db 3820 GCGCGCTTCCACTGAGCGGCTTCTGCGGAGGCGCGGCTGCGGAGGCGCTCGCGC 3879  
QY 1600 GCGCGCTCCCGGAGAACCGCGCACACGACATCGCAAGTACGGAGCAGACCTGTAC 1659  
Db 3880 GTCTGTGCTCTTCTCTCTGCTGCGCGGCTGCGGCGGCGGCTGCGGAGGCGGCTAC 3939  
QY 1660 GTGACCGGAGGACCTCGGAGCTCGCTGCGCGGCTGCTGCGCGGCTGCTGCGGAGCGCGC 1719  
Db 3940 GCGCGGCTACGCGCTTCTTCTGACGCGCTTCTGCGCGGCTGCGGCGGCGGCGCGC 3999  
QY 1720 GCGCGGCTACGCGCGGCGCGGATCGCTGACCTGCGCATCTCTGCTGACACCGCAGAGC 1779  
Db 4000 GTGACCTCGGTGGCGGCTTGGAGGCGGCTGCGGAGGCGGCTGCGGAGGCGGCTGCGG 4059  
QY 1780 GTCAGGTCTTGTCAACCGCGGCGCACACCGTGTCTTCCAGCAGGTCCACTTCCCGGAG 1839  
Db 4060 GAGCGGCTGCGCGGCTGCGGCTGCGCGGCTGCGCGGCTGCGCGGCGGCGGCTGCGG 4119  
QY 1840 GCGGACCGGGAATCTGCTCTTACACGACGCGCGCGCGCGGCTTCCAGCGG 1892  
Db 4120 GACACCGCGCTCGGCGCACGCGGACACCGCGCTGCTGACGATCGCGGAGCTGCTG 4172

RESULT 10  
US-09-861-289-32  
; Sequence 32, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

```
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861.289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match          3.9%; Score 78.2; DB 10; Length 11220;
Best Local Similarity 41.9%; Pred. No. 3.4e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCCTCGCTCCGCTCTACACATGACGCCCCCCCCCGGGTGGCTCTGGACACCCCA 528
Db 2740 GACTCCCCCGCGTCCAGGGTCTCCGTGAGGACTCTCTGGCGTACCGATCGACTGGA 2799
QY 529 CCGCGGTACACACACCGCGCTTACCAGCTGTACTACCTGCACTCCGACAGAACAC 588
Db 2800 CGCTTCGGGTCCGAGCGCTCCAGCGCGCGCGGGTGTCCGGCGGTGGCTCGTGTG 2859
QY 589 GGCCTCCGCGGTGGGACACGCGAGCACGACCGGCGGTGCTTACGCGACACCGCG 648
Db 2860 GTCCCGGAGGACCGTTCGCGCGGAGCGCGCGCGGTGCTCGCGCGCTGTCCGGCG 2919
QY 649 ACCGTGATCGCGCTCGGCGCGGACTTCCCGGTGCTGCTCGGGTTCGGCGGCGG 708
Db 2920 GCGGACCGCGTACAGCTGGAGCTGTCCCGGTGGCGGACCGCGGCTCGCGCGGAG 2979
QY 709 GCGAACACGCG ---AGGTGTGGCGCGCGCGGTGCTGCGCTCGCGACCGCGCGG 765
Db 2980 CTGGCGAGGCGCTTGGCGCGCGCGGTGGAGCGCTGCGAGCGGTCTCTGCTGCGG 3039
QY 766 GACGCGCTCGGAGTACGAGAGCAGTACCTCTACTGTGTGACGACGCGGGGTTAC 825
Db 3040 TGGGACGAGAGCGCACCGCGCGCGCGCGCGCGCTTACCGCGGGGACCGCGCGG 3099
QY 826 TTCACGCGCTCCCGACCGCGCTATCGTCAACACCGAGGTGCGCGCGCGCGCGG 885
Db 3100 CTCACCTGGTGGAGCGCTGGAGAGCGCGCGCGCTGCGCGCGCGCGCTGTGTGCG 3159
QY 886 GCGAGATCGAGACCGCGAGTGTTCGCGACCGCGCGAGATCCACTGGGACACCG 945
Db 3160 CACGCGCGGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3219
QY 946 GGAGATGGGTTCGCTCATCGGAGCTGCGGTACGCGCGGTTCACACTTCGCGGAC 1005
Db 3220 TGGGCGATGGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3279
QY 1006 CTGCGCGACTGGACACTTCGCGCGCACTTCGACTACCGACCGCGCGCTTCGGCGG 1065
Db 3280 CTGCGCGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3339
QY 1066 GAGTCCCGCGCGCTTTCGAGATCACCGCGAGCGGCGGACGCGCACTGGGTGCTG 1125
Db 3340 ACGGTGAGGACGAGTTCGCGGTACCGCGCTCCGGGTGCTCGCGCGCGCGCGG 3399
QY 1126 GCGAGCATGGAGCGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
Db 3400 GCGTCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3459
QY 1186 TGGGACGCGGAGCGTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1245
Db 3460 GTCACCGTGGCGAGGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3519
QY 1246 GACTGTACGCGCGCGTTCACCTGGCGCATGATCGACGCGCGCGCGCGCGCGCG 1305
Db 3520 GCGGACACCTCTCTCTCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 3579
```

## RESULT 11

```
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/860.846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match          3.9%; Score 78.2; DB 9; Length 36778;
Best Local Similarity 41.9%; Pred. No. 3e-09;
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;

QY 469 GCCTCGCTCCGCTCTACACATGACGCCCCCCCCCGGGTGGCTCTGGACACCCCA 528
Db 18427 GACTCCCCCGCGTCCAGGGTCTCGTGAGGACTCTCTGGCGTACCGCATCGACTG 18486
QY 529 GCGCGCGTCAACACCGCGCGCGCTTACCGCTGTACTACCTGCACTCCGACAGAAC 588
```

Db 18487 GCCTCGCGTCCGACGCGTCCGAGCGCGCGGCTGTCCGGGCGCTGGCTCGTCTG 18546  
QY 589 GGGCCCGGGGGTGGGACACGAGACGACGAGCGGCTTCCCTTACAGCACGACGCG 648  
Db 18547 GTCCCGAGGAGCGTTCCGCGGAGCGCGCGGCTGTCCGCGCTGTCCGCGCGCG 18606  
QY 649 ACCGTGATCCGCTGGGCGCGGATTTCCCGCTGTGGTTCGCGGTTCGTCGACAC 708  
Db 18607 GCGGACCCGTACAGCTGGACGTGTCCCGCTGGGCGACCGGCGGCTGTCCGCGG 18666  
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGGCGGCTGTCCGCTTCGAGACCGGAC 765  
Db 18667 CTGGGCGAGGCGCTGCGCGCGCGGCTGGAGCGCTGACGGGCTCTCTCGTGTCTG 18726  
QY 766 GAGCGCTGCGCAAGTACAGGAGTACCTTACTGTGACCGACGCGGCTTCACG 825  
Db 18727 TGGGACGAGCGCGACCCCGGCGCGCTTACCGGGGCGACCGGCGGCGAC 18786  
QY 826 TTCACCGCGCTGCGCGCGCGCTGAGGAGTACGCGGCTGCGCGCGCGCGCTGAC 18846  
Db 18847 CAGCGCGGCTGCTCGCTCGCGCGCGGCGCGACCGTACCTCCCGCGCGCGCGCTG 18906  
QY 946 GGAGATGGGTCTGCTCATCGGACGACTGCGGTAGCGCGCTTCTACACCTCGCGGA 1005  
Db 18907 TGGGGATGGCGGCTGCGCGCTTGGAGACCGCGGCTGGGCGCGCTGATCGAC 18966  
QY 1006 CTGCGGCTGAGACTTCGCGCGAATTCGACTACCGGAACCGCGCTCGCGCGCAT 1065  
Db 18967 CTGCGCTGAGCGCGCGCGCGCTTGGAGCGCATGACCGGTCTCTCGCGCGGT 19026  
QY 1065 GAGTGGCGCGCTGTCAGATACCGAGAGGAGCGGACGCGACTGGGTGCTCGCC 1125  
Db 19027 ACGGTTGAGGACGAGTCTCGGCTGCGCGCTTCTCGCGCGCGCTGCTCGCG 19086  
QY 1126 GCGAGATGAGCGCTTACGGCATCGCGCTTCCCATGACGTACGCTTACTGGACGAC 1185  
Db 19087 GCCTCCCTCCGCGCGCGCGCGCTTCCCGGTGGTGGAGCGCGCGCGCGCTG 19146  
QY 1186 TGGGAGCGGAGAGTTCCACGCGGAGCGCTTACCGCGAATGGTTCGACTGGGCTG 1245  
Db 19147 GTACCGGTGCGGAGAGCTTGGCGCGCGGAGCGCGCGCTGCGCGCGCGCG 19206  
QY 1245 GACTGTAGCGCGCTACCTTGGCCATCGATCGAGCGCGCGGAGACGCGCTCGCC 1305  
Db 19207 GCGGACACCTCTCTCTCACACCGCGCGCTTCCGCGAGCGGCGCGGCGCGCT 19266  
QY 1306 ATCGGTGGATGAACACTGGAAGTACGCGCGCGCGCGCTTCCCGACGCGCATCC 1365  
Db 19267 GTGCGCGCGGAGACTTCCGCGCTTCCCGGCTGTGTCGCGGAACTCGGCGCG 19326  
QY 1366 GGTACAGCGGAGAGTCTGATCTGCGGAGCTGCGGCTGCGCGGAGCG---CT 1419  
Db 19327 ACGGCGCGCTGTGACTTGGACCTTACGAGCGGAGGCGCGCGCGCTGCTCGCG 19386  
QY 1420 GCGGCTGTACACCTCTGTAGACACCGCGTGGAGCGCTGACGACTACGCTACGCG 1479  
Db 19387 GCGGTCTCGAGCGCGACCGCTGACGCGCGCTTCCACTTCCCGCGCGCGCTGAC 19446  
QY 1480 ACCACACACTTCCGCGCGCGCGTTCGAGCGGCGCGCTTCTGCGATGGAACGAG 1539  
Db 19447 GAGCGCTCGCGCGCGCGCGGAGCGGCTGCGCGCTGTGTCGCGGAGCGCGCG 19506  
QY 1540 GCATACGAGATCAGCTGACATCGCTTGGAGACCGCGGAGCGCGCTGCGGCTCT 1599  
Db 19507 GCGCGCTCCACTGAGCGCGCTTCTGCGGAGCGCGGCTGCGGAGCGCGCTCGCG 19566  
QY 1600 GCGCGCTCCCGCGCGCGCGCGCGGACGAGATCGGCAAGTACGAGGAGCGCTGAC 1659

Db 19567 GTCTCGGTCTCTTCTCTCTCGGTGCGCGGATCTGGGGCGGCGCGCGGTAC 19626  
QY 1660 GTCCGACGAGGACCTTCTGACCTTCCGCGGTACTCTCGCTCGCGCGCTTACTCGGAGCGCG 1719  
Db 19627 GCGCGGTTACGGCTTCTCTGACGCGCTTCCGCGGTACAGCGGCGCGCGCGCG 19686  
QY 1720 GCGCGCATCGACGCGCGCGGATCCGTGACCTTCCGCGGTCTCTGACACCGCGAG 1779  
Db 19687 GTGACCTCGGTGCGCGGTGGAGCGCTGGAGGCGAGCGCGTACCGAGGTGCGCGG 19746  
QY 1780 GTGAGGTCTTGTCAACCGCGCGCGACCGGTCTCTCCAGCAGGTCTCACTTCGCGGAG 1839  
Db 19747 GAGCGCTGCGCGCGCTGCGCGCTGCGCGCGCGCTGCGCGCGCGCGGTCTC 19806  
QY 1840 GCGCACACGGAATCTGCTCTACACCGCGCGCGCGCGCGCGGTCTCACTTCACCGG 1892  
Db 19807 GACACCGCGCTCGCGCGCGCGCGCGCGCGCGGTACGATCGCGCGGTGCGACTG 19859

RESULT 12  
US-09-861-289-5  
; Sequence 5, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-5

Query Match 3.9%; Score 78.2; DB 10; Length 36778;  
Best Local Similarity 41.9%; Pred. No. 3e-09;  
Matches 601; Conservative 0; Mismatches 823; Indels 9; Gaps 2;  
QY 469 GCTCGCTCCGTGCGCTTACCACATGAGCGCGCGCGCTGCTGCGACCGCGCGAA 528  
Db 18427 GACTCCCCCGCGTCCAGGCTCCGTGCGAGTCTCTGCGCTACCGATCGTGAAG 18486  
QY 529 GCGCGCGGTCAACCGCGCGCGCTTACAGCTGTACTACTGCTCCGCGCGCGCGCG 588  
Db 18487 GCGCTCGCGGTGCGCGCGCGCTTCCGCGCGCGCGCGCGCTGCGCGCGCTGCT 18546  
QY 589 GCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACGCGCGCGCGCG 648  
Db 18547 GTCCCGGAGGACCGTTCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCTGCTCG 18606  
QY 649 ACCGTGATGCGCGTGGCGCGCGCGCTTCCCGTGTGTGCGCGCGCGCGCTGCTCG 708  
Db 18607 GCGGACCGCGTACGATGAGCGTCCCGCGTGGCGCGCGCGCGCGCGCGCGCGAG 18666  
QY 709 GCGAACACGCG---AGGTTGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCG 765  
Db 18667 CTGGGCGAGGCGCTGCGCGCGCGCGGTGAGCGCGTGGAGCGCTCTCTCGCTGCT 18726  
QY 766 GAGCGCGTCCCAAGTACAGGAGCGTACCTTACTGTGTCGCGCGCGCGCGGTTCAG 825  
Db 18727 TGGGACGAGAGCGCGCGCGCGCGCGCGCGCGCTTACCGCGGCGCGCGCGCGCG 18786  
QY 826 TTCACCGCGCTGCGCGCGCGCGCTGAGGAGTACGCGCGCGCGCGCGCGCGCGCG 885  
Db 18787 CTCACCTGGTGCAGCGCTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18846







|    |       |  |       |
|----|-------|--|-------|
| Qy | 1186  | TGGGACGCGAGCAGTTCCACGCGCAGACACTACCCCGGCAATGGCTCGACTGGGGCTGG    | 1245  |
|    |       |  |       |
| Db | 20317 | GTCACCGGTCGCGAGGAGCCTTCGGCGCGCGAGCGCAGCCGCGTGGCCCGGAGCGC       | 20376 |
| Qy | 1246  | GACTGGTACGCGGCGCTCACCTGGCCATCTGACATCGACGCGCCCGAGACAACGCGCTCGCC | 1305  |
| Db | 20377 | GCCGGACACTCTCTCTCCACACACCCCTCCGGCAGCGAAGGCGCGAAGCACCTCC        | 20436 |
| Qy | 1306  | ATCGCGTGGATGAAACAATGGAACTAGCGCGCAGCGGAGCTCCCAACCGACGCATCCGAC   | 1365  |
|    |       |  |       |
| Db | 20437 | GGTGGCGCGAGGACTCCGGCTCGCCGGGCTCGTGCGCCAACTCGCGGAGACTGGGCGG     | 20496 |
| Qy | 1366  | GGCTACAAACGGCAGAACTCGATGCTCGCGAGCTGCGGCTCGCCCGACAGC-----CT     | 1419  |
| Db | 20497 | ACGGCCACCGCTGTGACTCGGACCTCAGCAGCGGAGGCGCGCGCGGCTGCTCGCC        | 20556 |
| Qy | 1420  | GGCGGCTGGTACACCCCTCTGAGCACCCCGCTGGCAGCGCTGAGCGAACTACGTCAACGCC  | 1479  |
| Db | 20557 | GGCGTCTCCGACGCGCACCCGCTCAGGCGCTCTCCACCTGCGCGCCACGGTCTCGACTCC   | 20616 |
| Qy | 1480  | ACCACCACATCCCGGACCGGACCGTGCAGGGCAGCGGCTCTGCGCATGGAACGGAGCG     | 1539  |
| Db | 20617 | GAGCGCTCGCGCGGACCGACGCGGAGCGGCTCGCCGCTGTGTCGAGCGGAAGCGCCACC    | 20676 |
| Qy | 1540  | GCATACGAGATCGAGCTCGACATCGCTGGGACACCGCAGCAGAACTCGGCTCATCGGTG    | 1599  |
| Db | 20677 | GCCCGGCTCCACTGGACCGGCTCTGCGGAGCGCGGGTGCAGGAGCGCTCCGCC          | 20736 |
| Qy | 1600  | GGCGCTCCCGGACGGAACCCGACACGAACTATCGGCAAGTACGAGCAGACTGTAC        | 1659  |
| Db | 20737 | GTCTGTGCTCTTCTCTCTCGGTCCGCGATCTGGGGCGCGCGGTCAAGGCGCGTAC        | 20796 |
| Qy | 1660  | GTCCACGAGGACCCCTCGGACTCGCGGGTACTCGCTCGCCCTACTCCGAGCGCGC        | 1719  |
| Db | 20797 | GCCCGGTACGGCTTCTCTGACGCGCTCGCCGGTCAGACCGGGCGGAGCGGCCACC        | 20856 |
| Qy | 1720  | GCCCCATCGACCCCGGCGCCGATCGGTGCACCTGGGCGATCTCTGTCGACACCCAGAGC    | 1779  |
| Db | 20857 | GTGACCTCGGTGGCTTGAGCGCCCTGGGAGGCGCGCGTCAACGAGGGTGCACCGGG       | 20916 |
| Qy | 1780  | GTCGAGGTCTTGTCAACGCGGGCCACACGCTGCTCTCCAGCAGGTCCACTTCGCGCGAG    | 1839  |
| Db | 20917 | GAGGGCTGCGCGGCTCTGGGCTCGGCGCTCTCGCCCGCGGAGGCGGCTACCGCGCTG      | 20976 |
| Qy | 1840  | GGCGACACGGGAATCTCTGCTCTATACGACGAGCGGCGGCCCGCACACTTCACCGG       | 1892  |
| Db | 20977 | GACACCGGCTCGGCGCACGGGACACCGCGCTCAGATCGCGGACGCTCGACTG           | 21029 |

| Query Match           | 3.9%  | Score 77.6        | DB 9      | Length 12441 |
|-----------------------|---|-------------------|-----------|--------------|
| Best Local Similarity | 42.9%   | Pred. No. 4.6e-09 |           |              |
| Matches 680           | Conservative 0  | Mismatches 879    | Indels 25 | Gaps 5       |
| Qy 367                | CGCGGCATCTACAGCGCGCGCGTCTCAGAGGAGCGCGCGGAGCACTCGCCCTGATC    | 426               |           |              |
| Db 11836              | CGGAGACACGGCAACTCTGACCCGACCAAGTGCGCGCGCGGTACACACCGCACTCG    | 11777             |           |              |
| Qy 427                | TTTCGGCGTGTCTGCCGCTGCAAGCCGGGATCCGCTCGCGGCTCGCTCGTCCGCTTC   | 486               |           |              |
| Db 11776              | GCCGTCTGGGTCTCCACTCTGGGGCGCCCTCGCGCGCGACCACTCGGTAAGTTC      | 11717             |           |              |
| Qy 487                | TACCAATGACGCCCGCCAGCGGTGGGTCTGCGACCCCAACGCCCGGTTCACCAACCAC  | 546               |           |              |
| Db 11716              | GCCGACGAGCACGGCTCGCGGTGTAATTCGACGCCGCGCAAGCCCTCGGCTGCGCGTC  | 11657             |           |              |
| Qy 547                | GCGGCTTACCAGTCTACTACTGATCTCGACACAGACAGGGCCCGCGGCTGGAC       | 606               |           |              |
| Db 11656              | GAGCGCGCGCGCGCGGACGCTCGCGACCGCGAGTGTTTACGTTCCAGCCCAACAG     | 11597             |           |              |
| Qy 607                | CACGGACGACGACGCGCGGTCCCTTTCAGCACACGACGACCGTGATCGCGTCGGG     | 666               |           |              |
| Db 11596              | GCGGTACAGCCTTCGAGGGGCGCGCTGTACCGAGAGCGCGCACTCGCGCCCGG       | 11537             |           |              |
| Qy 667                | CCCGACTTCCCGTGTGTCGGGTTCGGGTCTGTCGACCGCGCAACACGCAAGGTTTC    | 726               |           |              |
| Db 11536              | ATCCGCGCCTCCACAATCTCGGCTTCACCTGCGCGCGGACGCCCGCGCGGACCC      | 11477             |           |              |
| Qy 727                | GCGCGCGCGCGGTCTGCGGTCTGCGACACGACGCGACGAGGGGTCCGCAAGTACCAG   | 786               |           |              |
| Db 11476              | AACGCCAAGATGAGCGAGGCGCGCGCCCATG-----GGCTCACCTCCCTC          | 11429             |           |              |
| Qy 787                | GAGCAGTACCTCTACTGCTCGACGACGCGGGTTACGTTTCACGCGCTCGCCGACCCC   | 846               |           |              |
| Db 11428              | GACGGTTTCCGAGGTATCGACGGAACCGCGCAACACGCGCTCACCGGAGCAC        | 11369             |           |              |
| Qy 847                | GTATCTGTCACACACGACGCGTGGCGCGCCACACGCGCGCGAGATCGAAGACCGCAG   | 906               |           |              |
| Db 11368              | CTCGGGACCTCCCGGGCTCTCTGTCGCGACACGACGCGCCACGCGCTCAACACCCAC   | 11309             |           |              |
| Qy 907                | TGGTTCGCGACCCAGATCCATCTGGGACACGCGCGCGCGGAGATGGGTCTGGTCTATC  | 966               |           |              |
| Db 11308              | CAGTACGTGATCGTGTGAGATCGAGGAGCCACACCGCATCCACCGCACTCGTCAATG   | 11249             |           |              |
| Qy 967                | GGACGACTTCGGTACGCGCGTCTTACACCTTCGCGGAACCT-----CGCGACTGGACA  | 1020              |           |              |
| Db 11248              | GAGGTCTGAGGCGGAGGGCTGCACACCGCGCTACTTCTCGCGGGCTGCCACGAG      | 11189             |           |              |
| Qy 1021               | CTTCGCGCAACTTCGACTACCCGAACACGCGCTCGCGGCACTGAGTGCCTCGCACTG   | 1080              |           |              |
| Db 11188              | CTGAGCGGTACCGGGGACGCGACGCGCGCTCGCGCACACCGAAGCTCGCGCG        | 11129             |           |              |
| Qy 1081               | TTGAGATCACGACGACGACGACCGCCACTTGGGTCTCGCGCGCAGCATGACGCC      | 1140              |           |              |
| Db 11128              | GCGTGTCTCTCTCGCACGCGCACCGCCATCGCGGACGACGATCGCGGGTCTGCC      | 11069             |           |              |
| Qy 1141               | TACGCACTCGCCTCCCATCAGCTACGCTTCTGGACAGGCACTTGGAGCGGCGAGCAG   | 1200              |           |              |
| Db 11068              | GACCTGTCTGCTCTCTGCGGACCGCGCGCGGAGATGACCGCGCGCCAGCAACG       | 11009             |           |              |
| Qy 1201               | TTCCACGCGCAGCACTCACCCGCAATGGCTCGACTGGGCTGGGACTGGTACGCGGCC   | 1260              |           |              |
| Db 11008              | GCCCCGCC-CCGCTCGCGGCCCGCCAGACATCCAGCCCAAGTTGACGCTCCCCGATG   | 10950             |           |              |
| Qy 1261               | GTCATCTGGCCATGATGACGCGCGCGAGACCAAGCGCTCGCCATCGGCTGATGAC     | 1320              |           |              |
| Db 10949              | ACCGCCCCCGCCTTTCGCGCACGCGCGCGCGCAAGCTCGCGGCAACCCCGGAGCCGAT  | 10890             |           |              |
| Qy 1321               | AACGTGAAGTACGCGGACGCGAGTCCCCACGACGATCCGACGCTCAACAGGCGAG     | 1380              |           |              |
| Db 10889              | CTGGGGCGCGGCTCACGCGCTCGGCGAGACCTCTCGCGCGCGGCGCTCTGTCGCGCC   | 10830             |           |              |
| Qy 1381               | AACCTGATCTGTCG---CGAGCTCGGGCTCGCCCGACAGCCTGCGCGCTGGTACACCTC | 1437              |           |              |

```
Db 10829 GACGAGCGGGAACACCGCCCGCCACCTCGTCCGGCTCGCTACGGCAACAGC 10770
QY 1438 CTGAGCAGCCCGTGTGAGCGCTGACGAATACGTACCGCCACACACACATCTCCCGGAC 1497
Db 10769 CCGTTACCCCGCTGGAGAGCCCGCCAGACCTGGGGGTGACGGGACCCCTTCGG 10710
QY 1498 CGGACCGTGGAGGAGCGCGTCTCTGCTATGGAACGAGCGGCATACGAGATCGAGCTC 1557
Db 10709 CGCCTCCCTCGCCCTTTCGGGAGGTCCCGAGCTCCGACCGCGGTCCGAGACCGCC 10650
QY 1558 GACATCGCTGGGACACCGCGAGCAAGCTCGGCATCTCGGTGGGGCGCTCCCGCCAGGA 1617
Db 10649 GCGGGGGGTACTGGAAGAACACCTGCTCCCGCTCGAACAGCGGGGTCTTCGACGG 10590
QY 1618 ACCCGGCAACAGAA---CATCGGCAAGTAGGAGCAGACCTCTAGCTGACCGAGGACCC 1674
Db 10589 GCGCTCGCAGAGACCGCTTTCGGGTACAGCTGCGGCTTACCCGGCCGACCTCC 10530
QY 1675 TCGACCTCGCGGGTACTCGTCCGCCCTTACTCGGAGCGCGGCCCGCCATCGACCC 1734
Db 10529 ATGTTCCGCTGCACCTTCTGCTCGCTGTGACCGGCGCCGCTACGACCGCTCCGCCCTC 10470
QY 1735 GCGCGCCGATCGTCCACCTCGCATCTCTGTCGACACCCAGAGCGTTCAGTCTTCGTC 1794
Db 10469 GACGCGGCAAGCCCATGTTCCGGTGGGTATCGACGAGATACCGCGGGCAACCCCTCG 10410
QY 1795 AACGCGGCGCACACCGTCTCTCCACGAGGTCCACTTCGCGGAGGGGACACGGGAATC 1854
Db 10409 GCGATGTACTTCTCGCGGCTTGAGCGCTTACCAACCCGCGCTCGGAGCTGGCC 10350
QY 1855 TCGCTTACAGCGAGCGGGCGCCCGACACTTCACCGGATCGTCTCGCGGAGATGGC 1914
Db 10349 GCGCACGCGCACGAGCAGCGCTCGGGCCACCGTCTACACGAATCTCTTCGCTCACC 10290
QY 1915 CAGGCGATCTAGCGATGCACACC 1938
Db 10289 GAGCGACCTGGAGCGCAGCCC 10266
```

## RESULT 15

US-09-860-846-3/G

; Sequence 3, Application US/09860846

; Patent No. US20020164742A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/860,846

; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 3

; LENGTH: 13613

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-860-846-3

```
Query Match          3.9%; Score 77.6; DB 9; Length 13613;
Best Local Similarity 42.9%; Pred. No. 4.5e-09;
Matches 680; Conservative 0; Mismatches 879; Indels 25; Gaps 5;
```

```
QY 367 CGGGCATCTCACGGCGCGGTGTCTCCAGGAGCGCGCGGAGCACTCGCCCTGATC 426
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 12523 CGGACACCGCAACTCGACCCGACAGGTGGCGCGCGGTTCACACCCCGACCTCG 12464
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 427 TTCGGCGGTGTGTCCGCCCTGCAGCCCGGGCATCCGCTCGGGGTCTCGTCCCGTCC 486
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
```

```
Db 12463 GCCGTGTGGCGTCCCACTCTTGGGCGCGCCCTCGCGCGCGACGAGTCCGGAAGTTC 12404
QY 487 TACCACATACAGCCCGCCAGCGCTTGGCTCTCGCAGCCCGCAACGCGCGGTACACACCCAC 546
Db 12403 GCCGAGGAGCAGCGCTTGGCGTGTACTTCGACGCGCGCGACGCTCGGTTCGGGTGCGGGT 12344
QY 547 GCGGCTTACCAGCTGTACTTACCTGCTACCTCCGACCAAGAAACAGCGCGCGCGGTGGGAC 606
Db 12343 GACGCGCGCGCGCGCGGAGCTTCGCGGAGCGCGGAGGTTCAGGTTTCACGCCACCAAG 12284
QY 607 CACGCGAGCACACGAGCGGCTCGCCTTTCAGCACACCGGACCGGTGATCGCGTGGG 666
Db 12283 GCCGTCAAGCGCTTCAGGCGCGCGCGTCTGTCACGAGCGCGACCTCGCCGCGCGG 12224
QY 667 CCCGACTTCCCGCTGTGGTCCGGGTGCGCGGTCTGTCGCGCACCGCGCAACACGCGAGGTTC 726
Db 12223 ATCCGCGCGCTTCAACACTTTCGCTTCGACCTGCGCGGCGAGCCCGCGCGCGGAC 12164
QY 727 GCGCGCGCGCGGTCTGCGCTTCGCGACCCAGCGACCGGCGGTCTCGGAAGTACACG 786
Db 12163 AACGCGAAGATGAGGCGCGCGCGCGCGCATG-----GGCCTCACCTCCCTC 12116
QY 787 GAGCAGTACTCTACTTGTGTCGACGACGCGCGGTTCACGTTTACCGCGCTCGCCGACCC 846
Db 12115 GACGCTTTCGAGGTTCATGACCGGAACCGCGCGCAACGCGCGCTTACCGCGAGCAC 12056
QY 847 GTCATCTCAACACGAGCGGTGCGCGCGCACCGCGCGCGAGATCGAGAACGCGCGAG 906
Db 12055 CTCGCGGACCTTCCCGCGCTCTCTGTCGCGGACACGACCGCGCGCTCAACAACAC 11996
QY 907 TGGTTCCGGGACCCCAAGATTCATCTGGGACACCGCGCGGGGAGATGGTCTGCGTCTATC 966
Db 11995 CAGTACGTGATCGTGGAGATCGACGAGGCGCACCGCGCATCCACCGCGACCTCTCATG 11936
QY 967 GGACGACTCGGTAGCGCGCTTCTACACTCGCGCAACT-----GGCGACTGGACA 1020
Db 11935 GAGTCTCTAAGGCGCAAGCGGTGACACCCCGCGCTTCTTCGCGCGGTGCGCACGAG 11876
QY 1021 CTTTCGCGCACTTCGACTTACCCGAACACGCGCTTCGCGCGCATCGAGTTCGCCGACCTG 1080
Db 11875 CTGGAGCGGTACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11816
QY 1081 TTCGAGATCACCGGAGACGCGGACACCGCTCTGGTGTCTCGCGCGCGCGCGCGCGCGCG 1140
Db 11815 CGCGTGTCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11756
QY 1141 TACGCGATCGGCTCCCGATGACGTACGCTTCTGACAGGACCTTGGGACGCGCGAGCAG 1200
Db 11755 GACCTGTGCTCTCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11696
QY 1201 TTCCACGCGGAGCTTACCCCGCAATGGTCTGACTGGGGGTGGGACTTGGTACCGCGCG 1260
Db 11695 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11637
QY 1261 GTCACCTGGCCATGATCGACGCGCGCGGAGACCAAGCGCTCGCGATCGCGTGGATGAC 1320
Db 11636 ACCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11577
QY 1321 AACTGGAAGTACGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 11576 CTGGGGGCGCGGTCCACGCGCTCGCGCGAGACCTTCGCGCGCGCGCGCGCGCGCG 11517
QY 1381 AACTGATGTGCGG---CGAGCTGGGGTTCGCGCGAGACGCTTGGGGGTGTGTACACCTC 1437
Db 11516 GACGAGCGCGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11457
QY 1438 CTGAGCACCGCGTGGCGAGCGCTGACGAATACGTACCGCGCGCGCGCGCGCGCGCGCG 1497
Db 11456 CCGTTACCGCGCGTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11397
QY 1498 CGGACCGTGGAGCGCGCGCGCTTCTGCGATGAGAGCGGCGCATACGAGATCGAGCTC 1557
Db 11396 CGCCTCTCGCGCTTTCGCGGCGAGTTCGCGAGCTCCGCGCGCGCGCGCGCGCGCGCG 11337
```



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:32 ; Search time 34.9327 seconds  
(without alignments)  
1979.723 Million cell updates/sec

Title: US-09-868-328B-4  
Perfect score: 2837  
Sequence: 1 MTPAIRRAVLQAGAGALA.....LYTGGCPAHTGIVVREIGQ 519

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 2837  | 100.0       | 521    | 22 | Arthrobacter ureaf |
| 2          | 2352  | 82.9        | 517    | 20 | Arthrobacter nicot |
| 3          | 2352  | 82.9        | 517    | 23 | Levan fructotransf |
| 4          | 2265  | 79.8        | 484    | 20 | Arthrobacter nicot |
| 5          | 2265  | 79.8        | 485    | 23 | Levan fructotransf |
| 6          | 499.5 | 17.6        | 943    | 21 | Paenibacillus mace |
| 7          | 468.5 | 16.5        | 1277   | 21 | Paenibacillus pabu |
| 8          | 461   | 16.2        | 923    | 21 | Paenibacillus amyl |
| 9          | 427   | 15.1        | 516    | 21 | Novel recombinant  |
| 10         | 416.5 | 14.7        | 533    | 19 | C. utilis INV1 pro |

|    |       |      |      |    |                    |
|----|-------|------|------|----|--------------------|
| 11 | 389.5 | 13.7 | 490  | 16 | Penicillium purpur |
| 12 | 386   | 13.6 | 750  | 20 | Bacillus sp. L7 en |
| 13 | 386   | 13.6 | 750  | 20 | Bacillus L7 endo-1 |
| 14 | 366   | 12.9 | 556  | 15 | Inulinase preprote |
| 15 | 359.5 | 12.7 | 532  | 5  | Saccharomyces cere |
| 16 | 359.5 | 12.7 | 532  | 12 | Preinvertase. Sac  |
| 17 | 348   | 12.3 | 581  | 20 | Schizosaccharomyce |
| 18 | 348   | 12.3 | 581  | 20 | S. pombe invertase |
| 19 | 348   | 12.3 | 581  | 21 | Schizosaccharomyce |
| 20 | 292.5 | 10.3 | 671  | 23 | Fructosyl transfer |
| 21 | 277.5 | 9.8  | 636  | 16 | Tomato plant inver |
| 22 | 277.5 | 9.8  | 636  | 21 | Wild-type tomato i |
| 23 | 274.5 | 9.7  | 636  | 13 | Tomato vacuolar in |
| 24 | 274.5 | 9.7  | 636  | 14 | Tomato acid invert |
| 25 | 266   | 9.4  | 583  | 22 | Lycopersicon penne |
| 26 | 266   | 9.4  | 648  | 23 | Fructosyl transfer |
| 27 | 261.5 | 9.2  | 1116 | 20 | S. rochei strain E |
| 28 | 258.5 | 9.1  | 645  | 23 | Amino acid sequenc |
| 29 | 244.5 | 8.6  | 479  | 23 | Streptococcus poly |
| 30 | 243   | 8.6  | 670  | 21 | Corn invertase pro |
| 31 | 241   | 8.5  | 553  | 15 | Tomato acid invert |
| 32 | 241   | 8.5  | 553  | 16 | Tomato acid invert |
| 33 | 239.5 | 8.4  | 1487 | 16 | Bacillus circulans |
| 34 | 235.5 | 8.3  | 492  | 23 | Staphylococcus epi |
| 35 | 235.5 | 8.3  | 591  | 21 | Arabidopsis thalia |
| 36 | 231.5 | 8.2  | 587  | 23 | Arabidopsis thalia |
| 37 | 223.5 | 7.9  | 569  | 23 | Herbicidally activ |
| 38 | 222.5 | 7.8  | 554  | 23 | Herbicidally activ |
| 39 | 222   | 7.8  | 591  | 23 | Herbicidally activ |
| 40 | 220   | 7.8  | 581  | 23 | Herbicidally activ |
| 41 | 220   | 7.8  | 588  | 21 | Arabidopsis thalia |
| 42 | 205.5 | 7.2  | 630  | 17 | 1-Sucrose:sucrose  |
| 43 | 205.5 | 7.2  | 637  | 19 | Artichoke sucrose  |
| 44 | 202.5 | 7.1  | 429  | 22 | Corynebacterium th |
| 45 | 202.5 | 7.1  | 433  | 22 | C glutamicum prote |

ALIGNMENTS

RESULT 1  
AAB82301  
ID AAB82301 standard; Protein; 521 AA.  
AC AAB82301;  
DT 09-JUL-2001 (first entry)  
XX  
XX Arthrobacter ureafaciens levan fructotransferase.  
DE Arthrobacter ureafaciens levan fructotransferase.  
KW Levan fructotransferase; difructose dianhydride IV; sweetener.  
XX Arthrobacter ureafaciens.  
OS Arthrobacter ureafaciens.  
FH Key Location/Qualifiers  
FT Peptide 1..33  
FT Protein /label= Signal\_peptide  
FT /label= Mature\_protein  
XX WO200129185-A1.  
XX PD 26-APR-2001.  
XX PF 19-OCT-2000; 2000WO-KR01183.  
XX PR 19-OCT-1999; 99KR-0045302.  
XX PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.  
XX (REAL-) REALBIOTEC LTD.  
PI Rhee S, Song K, Kim C, Ryu E, Lee Y;  
XX

DR WPI; 2001-308483/32.  
XX N-PSDB; AAF30918, AAF30919.  
PT Producing difructose dianhydride IV from sucrose, involves reacting  
PT sugar solution in the presence of levansucrase to produce levan, and  
PT reacting levan solution in the presence of levan fructotransferase to  
PT produce DFA IV.  
XX Claim 2; Page 47; 72pp; English.  
XX The present sequence is that of Arthrobacter ureafaciens K2032  
CC levan fructotransferase (see AAB82301). The enzyme can be  
CC obtained by cultivation of claimed Escherichia coli JUD81  
CC (KCTC 0877BP), which carries claimed expression vector pUDFA81  
CC comprising the levan fructotransferase gene (see AAF30918). A  
CC claimed process for producing difructose dianhydride IV from  
CC sucrose comprises subjecting sugar solution to reaction at room  
CC temperature or lower in acidic buffer of pH 3.0-7.0 in the presence  
CC of a levansucrase derived from Zymomonas mobilis to produce levan,  
CC purifying the levan from the reaction solution, and subjecting it  
CC to reaction at 25-30 degree C for 3-10 hours in acidic buffer of pH  
CC 3.0-7.0 in the presence of levan fructotransferase, preferably  
CC obtained from E. coli JUD81. The product is useful as a low-calorie  
CC sweetener.  
XX  
SQ Sequence 521 AA;  
Query Match 100.0%; Score 2837; DB 22; Length 521;  
Best Local Similarity 100.0%; Pred. No. 1.2e-240;  
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTPAISRRVAVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVT 60  
DB 1 MTPAISRRVAVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVT 60  
QY 61 THGAYOLYLHSDQNGPGGDHASTTGDGVAFTHHGTVMPLRDPDFVWGSAAVVGTA 120  
DB 61 THGAYOLYLHSDQNGPGGDHASTTGDGVAFTHHGTVMPLRDPDFVWGSAAVVGTA 120  
QY 121 GFCAGAVVALATOPTDGVSRKYQOXYLYWSTDGFTTALPDPIVNTDGRAATTPAEIEN 180  
DB 121 GFCAGAVVALATOPTDGVSRKYQOXYLYWSTDGFTTALPDPIVNTDGRAATTPAEIEN 180  
QY 181 AEWFDPKIHWDTARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
DB 181 AEWFDPKIHWDTARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
QY 241 FEITADDGTRHWWLAASMDAYGIGLPMTYAYWTGTWGDGEQFHADDLTPOWLDGWDWYAA 300  
DB 241 FEITADDGTRHWWLAASMDAYGIGLPMTYAYWTGTWGDGEQFHADDLTPOWLDGWDWYAA 300  
QY 301 VTWPSIDAPETKRLATAMNNNKYAARDVPTDASDCYNGQNSIVRELRLARQPGGWYTL 360  
DB 301 VTWPSIDAPETKRLATAMNNNKYAARDVPTDASDCYNGQNSIVRELRLARQPGGWYTL 360  
QY 361 STPVAALTNYVTATTTLPDRTVGSAVLPWNGRAYEIELDIAWDTATNNGISVGRSPDGT 420  
DB 361 STPVAALTNYVTATTTLPDRTVGSAVLPWNGRAYEIELDIAWDTATNNGISVGRSPDGT 420  
QY 421 RHTNICKYGADLYVDRGSPDLGAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEVFVNA 480  
DB 421 RHTNICKYGADLYVDRGSPDLGAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEVFVNA 480  
QY 481 GHTVLSQQVHFAEGDGTGSLYTDGGPAHFTGIVVREIGQ 519  
DB 481 GHTVLSQQVHFAEGDGTGSLYTDGGPAHFTGIVVREIGQ 519  
RESULT 2  
AAY04105  
ID AAY04105 standard; Protein; 517 AA.  
XX  
AC AAY04105;

XX 10-JUN-1999 (first entry)  
XX Arthrobacter nicotinovorans levan fructotransferase prote.in #2.  
DE Arthrobacter nicotinovorans; levan fructotransferase.  
XX  
XX Arthrobacter nicotinovorans.  
OS  
XX Jp11069978-A.  
PN  
XX 16-MAR-1999.  
PD  
XX 28-AUG-1997; 97JP-0232421.  
PF  
XX 28-AUG-1997; 97JP-0232421.  
PR  
XX (NIOC ) NIPPON OIL CO LTD.  
PA  
XX WPI; 1999-247463/21.  
DR N-PSDB; AAX19827.  
XX  
PT Levan fructotransferase gene - for recombinant production of levan  
fructotransferase  
XX  
XX Claim 3; Page 8-9; 14pp; Japanese.  
XX The present sequence represents Arthrobacter nicotinovorans levan  
fructotransferase. The present invention also describes a method  
CC for the preparation of levan fructotransferase in which a transformant  
CC is cultured in a medium and levan fructotransferase is collected from  
CC the culture. The method can prepare levan fructotransferase in a  
CC large amount.  
XX  
SQ Sequence 517 AA;  
Query Match 82.9%; Score 2352; DB 20; Length 517;  
Best Local Similarity 81.3%; Pred. No. 5.1e-198;  
Matches 421; Conservative 41; Mismatches 52; Indels 4; Gaps 2;  
QY 1 MTPAISRRVAVLOGAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRPVT 60  
DB 1 MYDISRRALQAGAGALALFMSNAIPVAHAQA--SLRAIYHMTPPSGWLCDDPQRPVH 58  
QY 61 THGAYOLYLHSDQNGPGGDHASTTGDGVAFTHHGTVMPLRDPDFVWGSAAVVGTA 120  
DB 59 THGAYOLYLHSDQNGPGGDHASTTGDGVSYTHGVMVMPQDPFVWGSAAVVDTA 118  
QY 121 GFCAGAVVALATOPTDGVSRKYQOXYLYWSTDGFTTALPDPIVNTDGRAATTPAEIEN 180  
DB 119 GFCAGAVVALATOPTDGVSRKYQOXYLYWSTDGFTTALPDPIVNTDGRAATTPAEIEN 176  
QY 181 AEWFDPKIHWDTARGEWCVIGRLRYAAYFTSPNLRDWTLRNFDYPNHALGGIECPDL 240  
DB 177 AEWFDPKIHWDATRNENVCVIGRAYAAYFTSPNLRDQWQKSNFDYPNHALGGIECPDL 236  
QY 241 FEITADDGTRHWWLAASMDAYGIGLPMTYAYWTGTWGDGEQFHADDLTPOWLDGWDWYAA 300  
DB 237 FEMTAGDGTGRHWWVFGASMDAYSIGLPMTYAYWTGTWGNWTAFTADNLTPQWLDGWDWYAA 296  
QY 301 VTWPSIDAPETKRLATAMNNNKYAARDVPTDASDCYNGQNSIVRELRLARQPGGWYTL 360  
DB 297 VTWPAVEAPETKRLATAMNNNKYAARNYPTDASDCYNGQNSITRELRLERLGGWYTL 356  
QY 361 STPVAALTNYVTATTTLPDRTVGSAVLPWNGRAYEIELDIAWDTATNNGISVGRSPDGT 420  
DB 357 STPVPALSNYATSSITLPDRTVNGSVFLPWSGRAYELEIDISWDTAANVGVSGRSSDGS 416  
QY 421 RHTNICKYGADLYVDRGSPDLGAGYSLAPYSRAAAPIDPGARSVHLRLVDTOSVEVFVNA 480  
DB 417 RHTNICKYGADLYVDRASQSGYALAPYTRAAAPIDANARSVHLRIFVDTSQVEVFVNS 476  
QY 481 GHTVLSQQVHFAEGDGTGSLYTDGGPAHFTGIVVREIG 518



QY 278 GEOFHADDLTPQWLDGWDWYAAVTWPSIDAPETKRLAIAWMNNWXYAARDVPTDASGY 337  
 DB 241 GTAFIADNLTPQWLDGWDWYAAVTWPAVEAPETKRLAIAWMNNWXYAARNVPTDASGY 300  
 QY 338 NGONSIVRELRLAROPCGWYLLSTPVAALNTNYVTATTLPDRTVNGSAVLPHNGRAYEI 397  
 DB 301 NGONSITRELRLERQSGGWYLLSTPVPALSNYATSTTLPDRTVNGSEFVLPWNGRAYEL 360  
 QY 398 ELDIAMDTATNIGISVGRSPDGRHRTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPID 457  
 DB 361 ELDISWDTAANVCVSGRSDGSRHRTNIGKYGDELIVDRASSQSGYALAPYTRAAPID 420  
 QY 458 PGARSVHLRLVDTQSVFVFNAGHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREI 517  
 DB 421 ANARSVHLRIFVDTQSVFVFNAGHTVLSQQVHFAAGDTGISLYADGGPANFTGITIREF 480  
 QY 518 G 518  
 DB 481 G 481

RESULT 5  
 ABB10073  
 ID ABB10073 standard; protein; 485 AA.  
 AC ABB10073;  
 DT 01-JUL-2002 (first entry)  
 DE Levan fructotransferase amino acid sequence.  
 KW Levan fructotransferase; LFTase; difructose dianhydride; DFA IV;  
 KW Levan; enzyme.  
 OS Arthrobacter nicotinovorans.  
 XX  
 PN JP2002017366-A.  
 XX 22-JAN-2002.  
 XX  
 PF 06-JUL-2000; 2000JP-0205756.  
 XX  
 PR 06-JUL-2000; 2000JP-0205756.  
 XX  
 PA (NIPT ) NIPPON TENSAI SEITO KK.  
 PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 XX  
 DR WPI: 2002-287313/33.  
 DR N-PSDB; ABB10073.  
 XX  
 PT A new levan fructotransferase -  
 XX  
 PS Claim 2; Fig 1; 15pp; Japanese.  
 XX  
 CC The invention relates to a new levan fructotransferase (LFTase). LFTase  
 CC decomposes the beta-2,6-fructoside bond of levan in polyfructan to  
 CC produce difructose dianhydride IV (DFA IV). The LFTase of the invention  
 CC is used for the preparation of DFA IV. The current sequence represents  
 CC the levan fructotransferase amino acid sequence.  
 XX  
 SQ Sequence 485 AA;

Query Match 79.8%; Score 2265; DB 23; Length 485;  
 Best Local Similarity 83.0%; Pred. No. 2.1e-190;  
 Matches 399; Conservative 40; Mismatches 40; Indels 2; Gaps 1;

QY 38 SLRAVHYMTTPPSGLWCLDQRPVTHGAYQLYLHSDQNNPGGWDHASTDGVAFTHHGT 97  
 DB 4 SLRAVHYMTTPPSGLWCLDQRPVTHGAYQLYLHSDQNNPGGWDHASTDGVAFTHHGT 97  
 QY 98 VMLRDPFVWGSVAVGTANTAGFCAGAVVALATOPTDGVKRYQFOYLYWSTDGGFTFT 157  
 DB 98 VMLRDPFVWGSVAVGTANTAGFCAGAVVALATOPTDGVKRYQFOYLYWSTDGGFTFT 157

DB 64 VMPMQDPFVWGSVAVGTANTAGFCAGAVIALATOPTDG--KFQOYLYWSTDGGYSFT 121  
 QY 158 ALPDPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTRGIEWCVIGRLRYAAFTSPNLR 217  
 DB 122 ALPDPVIVNTDGRATTTPAEVENAEWFRDPKIHWDATRNWCVIGRARYAAFTSPNLR 181  
 QY 218 DWTLRNFDYPNHALGGIECPDLFEITADGTRHWVLAASMDAYGICLPMYAYWTGTWD 277  
 DB 182 DMQWKSNFDPNHALGGIECPDLFEMTAGDTRHWVFGASMDAYSICLPMYAYWTGSWN 241  
 QY 278 GEOFHADDLTPQWLDGWDWYAAVTWPSIDAPETKRLAIAWMNNWXYAARDVPTDASGY 337  
 DB 242 GTAFIADNLTPQWLDGWDWYAAVTWPAVEAPETKRLAIAWMNNWXYAARNVPTDASGY 301  
 QY 338 NGONSIVRELRLAROPCGWYLLSTPVAALNTNYVTATTLPDRTVNGSAVLPHNGRAYEI 397  
 DB 302 NGONSITRELRLERQSGGWYLLSTPVPALSNYATSTTLPDRTVNGSEFVLPWNGRAYEL 361  
 QY 398 ELDIAMDTATNIGISVGRSPDGRHRTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAAPID 457  
 DB 362 ELDISWDTAANVCVSGRSDGSRHRTNIGKYGDELIVDRASSQSGYALAPYTRAAPID 421  
 QY 458 PGARSVHLRLVDTQSVFVFNAGHTVLSQQVHFAEGDTGISLYTDGGAHFTGIVVREI 517  
 DB 422 ANARSVHLRIFVDTQSVFVFNAGHTVLSQQVHFAAGDTGISLYADGGPANFTGITIREF 481  
 QY 518 G 518  
 DB 482 G 482

RESULT 6  
 AAY91929  
 ID AAY91929 standard; Protein; 943 AA.  
 AC AAY91929;  
 XX  
 DT 19-JUL-2000 (first entry)  
 DE Paenibacillus macerans 2,6-beta-D-fructan hydrolase.  
 KW 2,6-beta-D-fructan hydrolase; anticarie; neuroprotective; biofilm;  
 KW hydrolyze slime; detergent; fructose production; sweetener.  
 XX  
 OS Paenibacillus macerans.  
 FH Key Location/Qualifiers  
 FT Protein 29..943  
 FT /label= mature\_protein  
 XX  
 PN WO200017331-A1.  
 XX  
 PD 30-MAR-2000.  
 XX  
 PF 17-SEP-1999; 99WO-DK00495.  
 XX  
 PR 18-SEP-1998; 98DK-0001173.  
 PR 09-DEC-1998; 98DK-0001623.  
 XX  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Moller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;  
 XX WPI: 2000-283564/24.  
 DR N-PSDB; AAA08535.  
 XX  
 PT New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful  
 PT e.g. for degrading microbial biofilm, hydrolyzing slime, production of  
 XX fructose oligomers and treatment of dental plaque  
 PS Claim 1; Page 119-120; 123pp; English.  
 XX  
 CC AAY91927-29 are polypeptides from Paenibacillus amylolyticus, P. papuli



CC and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase  
CC activity. These are preferred enzymes of the invention. The enzymes  
CC can be used to remove microbial biofilm from surfaces, particularly for  
CC reduction of biocorrosion, for treatment of dental plaque and for  
CC treatment of multiple sclerosis (eliminating biofilm from the lungs; to  
CC hydrolyze slime; in detergents and for production of fructose oligomers  
CC from levan or phleix (useful as sweeteners or for conversion to  
CC fructose). The nucleic acids that encode the enzymes are useful for  
CC recombinant production and as source of probes and primers for .  
CC identification and cloning of related fructan hydrolases (all claimed).  
xx  
SQ Sequence 943 AA;

**SQ** Sequence 943 AA;

Query Match 17.6%; Score 499.5; DB 21; Length 943;  
Best Local Similarity 30.8%; Pred. No. 1e-34;  
Matches 157; Conservative 67; Mismatches 223; Indels 63;

|    |     |   |     |
|----|-----|---|-----|
| Qy | 40  | RAYVHMTPPSGMLCDPQRPVTHGAYQLYLYLHSDQNNGPGGDHASTDGVAFTHHGTVM      | 99  |
| Db | 383 | RPDYHYSPARGSADNGLVYEGEYHLF---HOD---GCTWAHAYSTDLVHWKRLPIAL       | 436 |
| Qy | 100 | PLRPDPFVWSGSAVVGTAHTAGF-----GAGAVVALATOPTDGVKRYQBOYLWSTDDGS     | 154 |
| Db | 437 | PWDLGHVWSSAVADLHNASGLFADSGKGLIAYTYSNPDPNGNQIRIGILAYSKDRGR       | 496 |
| Qy | 155 | TFT-ALPDPVIVNTDGRAATPAEIEIAEM-PRDPKIHWDTARGEWCVCIGRLRYAAFYV     | 212 |
| Db | 497 | TWEYAEERPVTIENPGKQGGDP-----GGWDFRDPKVRDEEHNRMVMVSGGDHIREFT      | 551 |
| Qy | 213 | SPNLRDWTLRRNDYPNHALGGI-ECPDFELFETADD-GTRHVMVLAASMDAYGIGLPMTYA   | 270 |
| Db | 552 | STNLIDWTLTDSFGYAYVGGWGPCDLFQLAVDDTGEKKVLMISTCAN-----PNTQG       | 607 |
| Qy | 271 | YWTGWDGEGFHADDLTPOW-----LDGWDWYAAVTPSIDAPETKRLAIA               | 317 |
| Db | 608 | -----SAAEYFTEIGELTPEGKVFYNDNPAGKVLATDYKEYEYASMSFAGM--PDGRVMLA   | 659 |
| Qy | 318 | WMNNKYAARDVPTDASDGVNGONSIVRELRLAROPGGWYTLTSTPVAALTN-----YVTA    | 373 |
| Db | 660 | WMTNWDYPFA-FTT---EGWKGVLSIPRELTLOKTDKG-IRLAQPTIRELSRQQLIFA      | 714 |
| Qy | 374 | TTTLPORTVDG-----SAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGTTRHTNICKYGA | 430 |
| Db | 715 | AS---DRRVOADRENLLKGVSSGVYIEAEIETIPQASNVSEFGFRLREGAGKRTVVGYKT    | 771 |
| Qy | 431 | ---DLVYDRGPSDLAGYSLAPYSRAAPTDPCGARSVHLRILVDTOSVEFVNAGHTVLSO     | 487 |
| Db | 772 | KENEIVDRSLSGDGTGSERFTHLHOAPLOPONRRVKLRIFFDDSLSEFVGGDGRVVFSE     | 831 |
| Qy | 488 | QVHFAEGDGTGISLYTDGGAHFTGIYVUREI                                 | 517 |
| Db | 832 | VIFPDPAHREMSLFTVGGEVNVYSLKVHAL                                  | 861 |

RESULT 7  
AAY91928

| ID | AA | 91928 standard | Protein | 1277 AA |
|----|----|----------------|---------|---------|
| 1  | 1  | 1              | 1       | 1       |
| 2  | 2  | 2              | 2       | 2       |
| 3  | 3  | 3              | 3       | 3       |
| 4  | 4  | 4              | 4       | 4       |
| 5  | 5  | 5              | 5       | 5       |
| 6  | 6  | 6              | 6       | 6       |
| 7  | 7  | 7              | 7       | 7       |
| 8  | 8  | 8              | 8       | 8       |
| 9  | 9  | 9              | 9       | 9       |
| 10 | 10 | 10             | 10      | 10      |
| 11 | 11 | 11             | 11      | 11      |
| 12 | 12 | 12             | 12      | 12      |
| 13 | 13 | 13             | 13      | 13      |
| 14 | 14 | 14             | 14      | 14      |
| 15 | 15 | 15             | 15      | 15      |
| 16 | 16 | 16             | 16      | 16      |
| 17 | 17 | 17             | 17      | 17      |
| 18 | 18 | 18             | 18      | 18      |
| 19 | 19 | 19             | 19      | 19      |
| 20 | 20 | 20             | 20      | 20      |
| 21 | 21 | 21             | 21      | 21      |
| 22 | 22 | 22             | 22      | 22      |
| 23 | 23 | 23             | 23      | 23      |
| 24 | 24 | 24             | 24      | 24      |
| 25 | 25 | 25             | 25      | 25      |
| 26 | 26 | 26             | 26      | 26      |
| 27 | 27 | 27             | 27      | 27      |
| 28 | 28 | 28             | 28      | 28      |
| 29 | 29 | 29             | 29      | 29      |
| 30 | 30 | 30             | 30      | 30      |
| 31 | 31 | 31             | 31      | 31      |
| 32 | 32 | 32             | 32      | 32      |
| 33 | 33 | 33             | 33      | 33      |
| 34 | 34 | 34             | 34      | 34      |
| 35 | 35 | 35             | 35      | 35      |
| 36 | 36 | 36             | 36      | 36      |
| 37 | 37 | 37             | 37      | 37      |
| 38 | 38 | 38             | 38      | 38      |
| 39 | 39 | 39             | 39      | 39      |
| 40 | 40 | 40             | 40      | 40      |
| 41 | 41 | 41             | 41      | 41      |
| 42 | 42 | 42             | 42      | 42      |
| 43 | 43 | 43             | 43      | 43      |
| 44 | 44 | 44             | 44      | 44      |
| 45 | 45 | 45             | 45      | 45      |
| 46 | 46 | 46             | 46      | 46      |
| 47 | 47 | 47             | 47      | 47      |
| 48 | 48 | 48             | 48      | 48      |
| 49 | 49 | 49             | 49      | 49      |
| 50 | 50 | 50             | 50      | 50      |
| 51 | 51 | 51             | 51      | 51      |
| 52 | 52 | 52             | 52      | 52      |
| 53 | 53 | 53             | 53      | 53      |
| 54 | 54 | 54             | 54      | 54      |
| 55 | 55 | 55             | 55      | 55      |
| 56 | 56 | 56             | 56      | 56      |
| 57 | 57 | 57             | 57      | 57      |
| 58 | 58 | 58             | 58      | 58      |
| 59 | 59 | 59             | 59      | 59      |
| 60 | 60 | 60             | 60      | 60      |
| 61 | 61 | 61             | 61      | 61      |
| 62 | 62 | 62             | 62      | 62      |
| 63 | 63 | 63             | 63      | 63      |
| 64 | 64 | 64             | 64      | 64      |
| 65 | 65 | 65             | 65      | 65      |
| 66 | 66 | 66             | 66      | 66      |
| 67 | 67 | 67             | 67      | 67      |
| 68 | 68 | 68             | 68      | 68      |
| 69 | 69 | 69             | 69      | 69      |
| 70 | 70 | 70             | 70      | 70      |
| 71 | 71 | 71             | 71      | 71      |
| 72 | 72 | 72             | 72      | 72      |
| 73 | 73 | 73             | 73      | 73      |
| 74 | 74 | 74             | 74      | 74      |
| 75 | 75 | 75             | 75      | 75      |
| 76 | 76 | 76             | 76      | 76      |
| 77 | 77 | 77             | 77      | 77      |
| 78 | 78 | 78             | 78      | 78      |
| 79 | 79 | 79             | 79      | 79      |
| 80 | 80 | 80             | 80      | 80      |
| 81 | 81 | 81             | 81      | 81      |
| 82 | 82 | 82             | 82      | 82      |
| 83 | 83 | 83             | 83      | 83      |
| 84 | 84 | 84             | 84      | 84      |
| 85 | 85 | 85             | 85      | 85      |
| 86 | 86 | 86             | 86      | 86      |
| 87 | 87 | 87             | 87</    |         |

XX  
AC AAY91928;

DT 19-JUL-2000 (first entry)

DE *paenibacillus papuli* 2,6-beta-D-fructan hydrolase.

|    |  |
|----|--|
| XX | fructan hydrolase.   |
| XX |  |
| KW | 2,6-beta-D-fructan hydrolase; anticarie; neuroprotective; biofilm; |
| KW | hydrolyze slime; detergent; fructose production; sweetener.        |

XX  
OS  
XX  
OS

| XX | Key | Location/Qualifiers |
|----|-----|---------------------|
| FH |     |                     |

| FT | Protein | 25..1277 |
|----|---------|----------|
| FT | Protein | 25..1277 |

FT /label= mature\_protein

XX

PN WO200017331-A1.

XX  
PD 30-MAR-2000.

XX  
PF 17-SEP-1999; 99WO-DK00495.

18-SEP-1998; 98DK-0001173.  
XX

PR 09-DEC-1998; 98DK-0001623.  
XX

PA (NOVO ) NOVO-NORDISK AS.

XX  
PI Møller S, Johansen C, Schaefer T, Ostergaard PR, Hoeck LH;

XX  
DR WPI; 2000-283564/24.

DR N-PSDB; AAA08533.

XX New polypeptides with 2,6-beta-D-fructan hydrolase activity, useful  
PT e.g. for degrading microbial biofilm, hydrolyzing slime, production of  
PT fructose oligomers and treatment of dental plaque

XX  
PS Claim 1; Page 111-115; 123pp; English.

AAV91927-29 are polypeptides from *Paenibacillus amylolyticus*, P. pabuli  
and P. macerans, respectively, which have 2,6-beta-D-fructan hydrolase  
activity. These are preferred enzymes of the invention. The enzymes  
can be used to remove microbial biofilm from surfaces, particularly for  
reduction of biocorrosion, for treatment of dental plaque and for  
treatment of multiple sclerosis (eliminating biofilm from the lungs; to  
hydrolyze slime; in detergents and for production of fructose oligomers  
from levan or phlein (useful as sweeteners or for conversion to  
fructose). The nucleic acids that encode the enzymes are useful for  
recombinant production and as source of probes and primers for  
identification and cloning of related fructan hydrolases (all claimed).

|    |          |                    |
|----|----------|--------------------|
| CC |          | Identification and |
| XX |          |                    |
| SQ | Sequence | 1277 AA;           |

|                       |                  |                    |            |              |
|-----------------------|------------------|--------------------|------------|--------------|
| Query Match           | 16.5%            | Score 468.5;       | DB 21;     | Length 1277; |
| Best Local Similarity | 29.1%;           | Pred. No. 8.4e-32; |            |              |
| Matches 168;          | Conservative 74; | Mismatches 243;    | Indels 93; |              |

4 AISRRVLOGAGA-----GALALIFGGA VPPAA-----RASAPG----- 37

320 ASSNRKVPSPQGRHHIEVIASSGLIQVYVDGTPAAVEVTDKSYAKGNAGLVVQOGMAY 379

QV 38 -----SLRAVYHMTPPSGWCLDPORPVTHGAYQLYYLHSDQNNRGGWD 82

380 FQDIUNTEESMYKENRYROYHYSPURGSASDNGLVYEGEYHLF---HQD-----GGTWA 433

QV 83 HASTTDCGVAFTHHGCTVMPRLRPDPFVWVGSSAVVCTANTAGE-----GAGAVVALATQPTDQ 137

434 HAVSSDLINWKRIPALPNWDQGHVWVGSSAIADLNNSAGLETTDSGGKGLIAVYTSYHPDK 493

QY 138 VRKYQEQLYXWSTGGGFTT-ALPDPVIVNTGRAATTAEIENAEW-FRDPKIHWD TAR 195

Db 494 PGGNORIGLAYSTDQGRNQWYAKERIVIDNPGKNGDDP-----GSWDFRDPKVVREDH 548

QY 196 GEWCVIGRLRYAAFYTSNLRDWTLRNFDPNHALGGI-ECPDFEITAD-DGTRHWV 253

Db 54.9 NRWVMVSGGDHRRFTSTNLLDWTLTDFEGYDGVYRGVWCEPDLQLPVDCTGQRKW 608

QV 254 LAASNDAYGIGLPMTYA----YWTG--TWGGEFHADDITPQWL--DWGWDNYAAVTWPS 305

Db 609 LLIS<sup>TM</sup>GAN----PKTQGSDAEYFVGOLTAG-KFLNDHPAGQVLR<sup>TM</sup>TDYCKEFTVASMFSAN 663

QY 306 IDAPETKRLAIAMNNWVYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLILSTPVA 365

Db 664 M-PNQKVMLAWMTNWDYPF-EFTSS---WKGLTIPREVSRLTTDEG-VRLVQTPIT 716

366 ALTNVVTATTTTPDRTVDCSAVLPHNG---RAYEIELDI---AWDTATNVGISVGRSPDG 419 QY

Db 717 ELQKLRHNLYSAQOMTVGPKSKNPLEGLTAGAYEIAEAEVIPANSSVTEFGFOL-RQREG 775

QY 420 TRHTNIGKYGAD---LYVDRGSPDLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTSQVEV 476



```
XX SQ Sequence 516 AA;
Query Match 15.1%; Score 427; DB 21; Length 516;
Best Local Similarity 27.2%; Pred. No. 9.9e-29;
Matches 145; Conservative 92; Mismatches 221; Indels 76; Gaps 21;

Qy 24 GGAVPAAARASAPGSLRAYVHTPPSGWLCDPQRPVTTTHGAYQLYVLSHDQNNPGG--W 81
Db 16 GLTLPSQAQSN---DYRSPSYTPDOYWMNEPGLIKIGSTWHLFFQHNPNTANVGNICW 72

Qy 82 DHASTTDCVAFTHHCTVMPRLPDPFVWSGSAVGTANTAGGAGA---VVALATOPTGV 138
Db 73 GHATSTDLHMAHKPTADENGVEAFTGTAYDPNNTSGLDNSANPPYLAFTGYTTS- 131

Qy 139 RYQOQYLYWSDGFTFTALPDVIVNTDGRAATTAEIENAEFRDPKIHWDTARGEW 198
Db 132 SOTQDORLAFSDNGATWTKFQGNPIISTSQEA---PHDITGGLESRDPKVFHFRQSNW 188

Qy 199 VCVI--GRLRYAAFTSPNLRDWTLRRNFDYPNHALGFI-----ECPDLFEITADGT 249
Db 189 IMVLAHGGQDKLSFWTSADTINWTQS--DLKSTSLNGLSSDITGWEVPMFELPV-EGT 245

Qy 250 RH--WVL-----AASMDAYGIGLPMYAYWTGTWGEQFHAD--DLTPQWLWDGWDYAAV 301
Db 246 BETTWVMMTPAEGSPAGGNGVLAI---TGSFQKSTADPVPDASTMWLDNGRDFDGL 301

Qy 302 TWPSIDAPETKRLATAMNNKYYAARDVPTDASDGYNGONSIVRELRLAROPGGWYTLIS 361
Db 302 SWNVNPSADGRRIITAAVWNSY---GSPNPTTT---WKMLSPFRTLSL-KKVGTOOHRVQ 354

Qy 362 TPVAALTNVYVATTPLPRTVDGSAVLPWNGRAYEIELDIAW--DTATNVGISVGRSPDG 419
Db 355 QPITELDTISTSLQILANOTITPGQTLSSIRGTALDVRVAFYPDAGSVLSLVRK--G 411

Qy 420 TRHNTIGXY---GADLYVDRGPSDLAGYSLAPYSRAAIPIDPGARSVH-----464
Db 412 ASEQTVIKYTSODAPLSVDRTESGDISY-----DPAAGGVHTAKLEEDGTGLV 459

Qy 465 -LRILVDFQSVVEFVNAGHTVLSQVHFAEGDTGISLTDGGPAHFTGIVVREI 517
Db 460 SIRLVDFQSVVEFGGQGEAVISDLIFFSDSDGLALEVTCGNAVLSQVDVRSV 513

RESULT 10
AAW44863
ID AAW44863 standard; Protein; 533 AA.
XX
AC AAW44863;
XX
29-SEP-1998 (first entry)
XX
C. utilis INV1 protein.
XX
UR3; HI3; INV1; orotidine 5'-monophosphate decarboxylase; enzyme;
KW biosynthesis; ribonucleotide; uracil; histidine; auxotroph; invertase;
KW imidazole-glycerol-phosphate dehydratase; beta-fructofuranosidase;
KW genetic marker; plasmid; transformation.
XX
OS Candida utilis.
XX
WO9814600-A1.
XX
09-APR-1998.
XX
03-OCT-1997; 97WO-CU000005.
XX
03-OCT-1996; 96CU-0000082.
XX
(INGG-) CENT ING GENETICA & BIOTECNOLOGIA.
XX
Besabe Tuero L, Chavez Espinoza FP, Delgado Boada JM;
PI Gonzalez Martinez ME, Paifer Reyes E, Rivero Baeza T;
XX
```

```
PI Rodriguez Menocal L;
XX
DR WPI; 1998-240099/21.
XX N-PSDB; AAV19464.
XX
PT New Candida utilis auxotrophic mutants and DNA markers - useful in
PT transformation systems to allow selection of transformants, e.g. for
PT heterologous protein production
XX
PS Disclosure; Page 29-31; 63pp; Spanish.
XX
CC This is the amino acid sequence of the Candida utilis INV1 enzyme
CC invertase (beta-fructofuranosidase). The invention relates to novel
CC C. utilis strains lacking one of the UR3 (AAV19462), HI3 (AAV19463), or
CC INV1 genes resulting in auxotrophic mutants. The genes can be used
CC as markers in plasmids for transforming C. utilis. This organism has
CC not been used as a host for transformation due to difficulties in
CC obtaining reliable markers. Straightforward selection of transformants
CC using the new auxotrophic mutants as host cells is now possible.
XX
SQ Sequence 533 AA;
Query Match 14.7%; Score 416.5; DB 19; Length 533;
Best Local Similarity 25.9%; Pred. No. 8.7e-28;
Matches 132; Conservative 87; Mismatches 182; Indels 109; Gaps 22;

Qy 40 RAVYHMTPPSGWLCDPQRPV---TTHGAYQLYVLSHDQN---NGPGWDHASTTDCGVAFT 94
Db 30 RPLVHLTPPVGMNDPNGLFYDSSSTHYVYQYNNPNTINGLPLYWGHATSDLLTWDH 89

Qy 95 HG--TYMPLRDPFVWSGSAVGTANTAGGAGA-----VVALATOPTDGVRYQEQYLY 147
Db 90 HAPALGPNDEBGIYSGSIVDIDYNTSGFFDSTREQRIVAIYTNLPLDV---ETQDIA 146

Qy 148 WSDGCGFTTALPDVIVNTDGRAATTAEIENAEFRDPKIHWDTARGEWVCVIGRLR- 206
Db 147 YSDGCGYTFEKYENNPVIDV-----NSTQFRDPKVIWYEETEQWMTVAKSQE 194

Qy 207 -YAAFTSPNLRDWTLRRNFDYPNHALGIECPDLFEITAD-----DGRHVLAAASMDA 260
Db 195 YKIQIYTDNLKDWLSASNFSTKGVGYQYECPLFEATIENTPKSGDPEKKWVWLAINP 254

Qy 261 YGIGLPM--TYAYWTGTWGEQFHADDLTPQWLWDGWDYAAVTPWPSIDAPETKRLAIA 317
Db 255 ---GSPLGGSINEXEYVGDFNGTEFIPDDDDATRFMDTKDFYAFQAF--FNAPENRSIGA 309

Qy 318 WMNNKYYAARDVPTDASDGYNGONSIVRE-----LRLAROPGGWYTLSTTP 363
Db 310 WSSNMQY--SNQVPD--PDGYRSSMSSIREYTLRYVYSTNPESQQLILCQK---FEVNETD 363

Qy 364 VAALTNVYVATTPLPRTVD---GSAVLPWNGRAYEIELDIAWDTATNVGISVGRSPDGT 420
Db 364 LKVVVEYKVSNSL---TVDHFTGSSFANSNTTGL---LDF-----NMTFTVNGTTDVT 411

Qy 421 RHTNIGKYGADLYVDRGPSDLA---GYSL-----APYSR 451
Db 412 QKDSV---TFELRIKSNQSDAIALGYDYNNEQFYINRATESYFQRTNQFOERWSTVYQ 468

Qy 452 AAAPIDPGARSVHLRLIVDTQSVEFVNAG 481
Db 469 PLTITESGDKQYQLYGLVDNILELYFNDG 498

RESULT 11
AAR72566
ID AAR72566 standard; Protein; 490 AA.
XX
AC AAR72566;
XX
23-NOV-1995 (first entry)
XX
Penicillium purpurogenum endo-inulinase.
XX
```





CC the S. cerevisiae SUC2, GAL4 and GAL1 genes. The inulinase gene appears  
 CC to be regulated solely by glucose repression which allows the  
 CC construction of a strong, non-repressible promoter by exchanging the  
 CC MIG1 DNA binding site in the inulinase promoter. The regulatory region  
 CC of the inulinase gene comprises a promoter, an upstream activating or  
 CC repressing sequence (UAS or URS), a terminator as a regulatory region,  
 CC and a sequence encoding a secretory signal necessary for secreting a  
 CC gene product from a yeast. This regulatory region may be used to  
 CC prepare transformed yeasts for the production of desired gene products,  
 CC eg. protein or metabolites.

XX Sequence 556 AA;

Query Match 12.9%; Score 366; DB 15; Length 556;  
 Best Local Similarity 27.0%; Pred. No. 2.5e-23;  
 Matches 132; Conservative 72; Mismatches 181; Indels 104; Gaps 24;

QY 40 RAYVHTPPSGNLCDPQRPVTHGA-----YQLYLHSDQN---NGPGGWDHASTTD 88  
 DB 39 RPSVHTPSHGWNNDP-----NLWDYDAKEEDHLYQYNPAATIGWTPLYGHAVSKD 92  
 QY 89 GVAFTHHGTVM-PLRPDPFVWGSAGVGTANTAGF-----GAGAVVALATOPTDGV 138  
 DB 93 L7SWTDYGLSLGPGSDDAGAFSGSMVIDYNTSGFNSVDPRQRAVAVWTLSKGPS--- 149  
 QY 139 RYQOYLWSTGGFTFTALPDPVIVNDGRAATTPAEIENAEFRDPKIHWDIAR--- 195  
 DB 150 ---QAQHSYSLDGGYTFQYSDNAVLDT-----NSSNFRDPKVFWEHGE 194  
 QY 196 -GEWVCVIGRLAY--AAFTYSPNLRDWTLLRRNFDYPNHALGIECPDL---FEITADGG 248  
 DB 195 DGRWIMAEOSQVFSVLFYSSPNLKNWLTESNFTHHGWTGTQYCEPGLVKVPYDSVADSS 254  
 QY 249 TRH-----WVLAASMDAYG-IGLPMYAYWTGTWGEQFHADDLTLPQWLDGWDWA 299  
 DB 255 SNSSDKSPDSANVLFVSNIPGGLGSSVT-QYFVGDFNGTHETPIDDDQTRFLDMGDIYA 313  
 QY 300 AVTWSIDAPEYKRL-AIAMNNWYKAAARDVPTDASDGYNGQNSIVRELRL---ARQGG 355  
 DB 314 LQTF--FNTPNKDYVGTAWASNWOY-AQAAPTDP---WRSSMSLVROFTLKDFTNPN 367  
 QY 356 WYTL-STEPVAALTNVYATTTLPRTVDGSAVLWPNGRAYEILDIAWDATNVGISVG 414  
 DB 368 ADVVLNSQPV---LNTDALRKNNGTYSITNYVTSENGK--KIKLD----- 408  
 QY 415 RSPDGRTHNIGYGADLYVDRGSPDLGAGYSLAPYSRAAAPIDPGARSVHLRLVDTQSV 474  
 DB 409 -NPGSLEPHLE-----YVFNQSPDKSNVFNADLSLYFKGNDD--NEYRLGYETNGG 459  
 QY 475 EYFVNAGHT 483  
 DB 460 AFFLDRGHT 468

RESULT 15  
 AAP40020  
 ID AAP40020 standard; Protein; 532 AA.

XX AAP40020;

XX 30-NOV-1991 (first entry)

XX Saccharomyces cerevisiae invertase.

XX Signal sequence; invertase; vector; leukocyte interferon.

XX Saccharomyces cerevisiae.

XX EP127304-A.

XX 05-DEC-1984.

XX 24-APR-1984; 84EP-0302722.

XX

PR 25-APR-1983; 83US-0488337.

XX (GETH ) GENENTECH INC.

XX Chang CN, Matteucci MD, Hitzeman RA;

XX WPI; 1984-301996/49.

DR N-PSDB; AAM40009.

XX Yeast expression vehicle - contains DNA sequence of yeast promoter

PT linked to DNA sequence encoding homologous protein.

XX Disclosure; Fig. 13; 53pp; English.

XX The DNA encoding the signal peptide of *S. cerevisiae* invertase may  
 CC be used in a new yeast expression vector for the expression of, esp.  
 CC human leukocyte interferon-alpha. The expression vector comprises a  
 CC yeast promoter operably linked to the signal sequence. This is attached  
 CC to the coding sequence of the protein. Yeasts transformed with this  
 CC vector produce and secrete protein normally heterologous to them.  
 CC The protein can be recovered from the medium free of presequences  
 CC etc. Human leukocyte interferon can be used as an antiviral or  
 CC antitumour agent. The promoter is esp. the invertase promoter or  
 CC the yeast 3-phosphoglycerate kinase promoter.

XX Sequence 532 AA;

Query Match 12.7%; Score 359.5; DB 5; Length 532;

Best Local Similarity 26.0%; Pred. No. 8.9e-23;

Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;

QY 19 LALIFGAVPPAARASAPGSLRAVYHMTTPSGWLCDP-----QRPVTHGAYQLYLLHS 72  
 DB 7 LELLAGFAAKISASMTNETSDRPLVHFETPKGMNDPGLWYDEKDAWH---LYFOYN 62

QY 73 DON---NGPGGHDHASTTDGVAFTTHG-TVMPLRDPFVWGSAGVGTANTAGFGAGAV- 127

DB 63 PNDTVMGTPLEWGHATSDDLTNWEDQPIAIAPIKNDGAFSGSMVVDYNTSGFFNDTID 122

QY 128 -----VALATQPTDGVKRYQOYLWSTGGFTFTAL-PDPVIVNDGRAATTPAEIENA 181  
 DB 123 PRQRCVAINTYTFP---ESEEQYISYSLDGGYTFTEYQKNPVL-----ANS 166

QY 182 ENFRDPKHWDTARGEWCVICGLRL--YAAFTYSPNLRDWTLLRRNFDYPNHALGG--IEC 237  
 DB 167 TQFRDPKVFWEYPSQKWIIMTAASKQDYKIEIYSSDDLKSKWLESAP--ANEGLFGYQEC 224

QY 238 PDLFEITA--DDGTRHWVLAASMDAYGICLPMTYA---YWTGTWGEQFHADDLTQWLD 292  
 DB 225 PGLIEVPTQDPSKSYWVNFISINP---CAPAGGSFNQYFVGSENGTHEAFEDNOSRVVD 281

QY 293 WGDWYAAVTPSIDAPETKRLAIAAMNNWYKAAARDVPTDASDGYNGQNSIVRELRL--- 349  
 DB 282 FGKDYALOTFTFTDPTGSAIGIAWASNWEYSA-FVPTNP---WRSSMSLVKRFSLNTE 337

QY 350 ---AROPGGWYTLSTPVAALTN-----YVTATTLTPDRT--VDGSAVLWPNGRAYEIEL 399  
 DB 338 YOANPETELINLKAEPILNISNAGPWSRPAATNTTLTKANSYND-----LSNSTGTLEFEL 393

QY 400 DIAWDATNVGISVG-----RSPDGRTHNIGYGADLYVDRGSPDLGAGYSLAPY 449  
 DB 394 VYAVNTQTQISKSVFADLSLWPKGLEDPPEYLRMGFEVSASSFFLDRGSKVKFKENPY 453

QY 450 --SRAAAPIDPGARSVHLRI-----LVDTPQSVFVFNAGHTVLSQOVHFAEGD----- 495  
 DB 454 FTNRMSVNNQPEKSENDSLSYKVGILLDQNILELFENDGVVSTNTYFTTGNALGSVM 513

QY 496 -TGIS--LYTD 503

DB 514 TTGVNDNLFYID 524

Search completed: May 6, 2003, 12:32:14  
Job time : 36.9327 secs

---



**THIS PAGE BLANK (USPTO)**

| Result No. | Query |       |      | Length | DB     | ID                                       | Description |
|------------|-------|-------|------|--------|--------|--|-------------|
|            | Score | Match |      |        |        |  |             |
| 1          | 2352  | 82.9  | 517  | 2      | JC5875 | levansucrase (EC 3.2.1.1)                |             |
| 2          | 944   | 33.3  | 516  | 2      | E70035 | levanase homology                        |             |
| 3          | 880.5 | 31.0  | 514  | 2      | G97118 | levanase [imported]                      |             |
| 4          | 562.5 | 19.8  | 512  | 2      | A5206  | sucrose/fructan hydrolase                |             |
| 5          | 556   | 19.6  | 677  | 2      | A27286 | levanase (EC 3.2.1.1)                    |             |
| 6          | 515   | 18.2  | 622  | 2      | A36915 | fructanase - Bacteroides                 |             |
| 7          | 467.5 | 16.5  | 943  | 2      | JC4081 | sucrose/fructanase                       |             |
| 8          | 434   | 15.3  | 516  | 2      | JE0301 | inulinase (EC 3.2.1.1)                   |             |
| 9          | 425   | 15.0  | 515  | 2      | JC7533 | inulinase (EC 3.2.1.1)                   |             |
| 10         | 401.5 | 14.2  | 1142 | 2      | CG7080 | levanase/invertase                       |             |
| 11         | 389.5 | 13.7  | 515  | 2      | JC5458 | inulinase (EC 3.2.1.1)                   |             |
| 12         | 382.5 | 13.5  | 1423 | 2      | A49206 | exo-beta-D-fructosyltransferase          |             |
| 13         | 379   | 13.4  | 550  | 2      | S65753 | beta-fructofuranosyltransferase          |             |
| 14         | 369.5 | 13.0  | 532  | 2      | S27372 | beta-fructofuranosyltransferase          |             |
| 15         | 366.5 | 12.9  | 532  | 2      | S27373 | beta-fructofuranosyltransferase          |             |
| 16         | 366   | 12.9  | 556  | 1      | S13130 | inulinase (EC 3.2.1.1)                   |             |
| 17         | 365.5 | 12.9  | 555  | 1      | S17502 | inulinase (EC 3.2.1.1)                   |             |
| 18         | 359.5 | 12.7  | 532  | 1      | IFBY   | beta-fructofuranosyltransferase          |             |
| 19         | 355   | 12.5  | 477  | 2      | C91034 | sucrose-6-phosphatase                    |             |
| 20         | 354   | 12.5  | 477  | 2      | S52162 | sucrose hydrolase                        |             |
| 21         | 348   | 12.3  | 477  | 2      | D58178 | sucrose hydrolase                        |             |
| 22         | 348   | 12.3  | 581  | 2      | JE0102 | invertase - fission yeast                |             |
| 23         | 330   | 11.6  | 476  | 1      | C43717 | beta-fructofuranosyltransferase          |             |
| 24         | 320   | 11.3  | 533  | 2      | S13528 | beta-fructofuranosyltransferase          |             |
| 25         | 318   | 11.2  | 490  | 2      | T43149 | probable beta-fructofuranosyltransferase |             |
| 26         | 313.5 | 11.1  | 453  | 2      | T39155 | glycosyl hydrolase                       |             |
| 27         | 312   | 11.0  | 590  | 2      | T02096 | probable beta-fructofuranosyltransferase |             |
| 28         | 297.5 | 10.5  | 581  | 2      | T12095 | beta-fructofuranosyltransferase          |             |
| 29         | 283.5 | 10.0  | 580  | 2      | S49308 | beta-fructofuranosyltransferase          |             |

QY 301 VTWPSIDAPETKRLATAMNNKYYAARDVPTDASDNGNSIVRELRLARQCGGWTLL 360  
Db 297 VTPVAEAPETKRLATAMNNKYYAARNVPTDASDNGNSIVRELRLARQCGGWTLL 356  
QY 361 STPVAALTNVYATTTLPDRTPDGSVLPWNGRAYEIELDIAMDATNVGSGSPDGT 420  
Db 357 STVPALSNVYATSTTLPDRTPVNGSVLPWNGRAYEIELDIAMDATNVGSGSDGS 416  
QY 421 RHTNICKYGADLYDRGPDGLAGYSLAPYSRAAAPIDPGARSVHLRLVLTQSVYEVNA 480  
Db 417 RHTNICKYGDELVDYDRASSBSQYALAPYTRAAPIDANARSVHLRIEVDQTSVEFVNS 476  
QY 481 GHTVLSQVHFACDYGISLYTDGGAHFTGVVREIG 518  
Db 477 GHTVVSQVHFAGADYGISLYADGGPANFTGIREFG 514

## RESULT 2.

E70035

levanase homolog yveB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: E70035

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.: Ehrlich, S.D.; Emerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997A:Authors: Hentz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim  
wood, C.R.; Hentz, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.  
Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis, S.; Lauber,  
A:Authors: Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigu  
K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwol  
Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleid  
A:Authors: Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro  
amakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol  
Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa,  
A:Authors: Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:198044033; PMID:9384377

A:Accession: E70035

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-516 &lt;KUN&gt;

A:Cross-references: GB:290121; GB:AL009126; NID:92635827; PIDN:CABL5451.1; PID:g2635959

A:Experimental source: strain 168

C:Genetics:

C:Gene: yveB

C:Superfamily: Penicillium purporogenum inulinase

Query Match 33.3%; Score 944; DB 2; Length 516;  
Best Local Similarity 41.8%; Pred. No. 4e-62;  
Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAVYHMTTPSGWLCQDPQPVTHGAYQLYLHS -DQNGPG-GWDHASTDGVAFTHHGT 97  
Db 44 RAYHETFDKKNDDPKPIYDPKTHYLYNDRYPKNGTEWRHVSDDLHVHTDEGV 103  
QY 98 VMP--LRPDPFVMSGSAVVGTAAGAGAVVALATQPTDGVKRYEQOYLWSTDGGFT 155  
Db 104 AIPKYNPDGDIWGSVVVDKENTAGFGKNAIVATQPSAKDKK-QEQLWYSTDKGS 162  
QY 156 F-----TALPDPIVNTDGRAATTPAEIENAEWFRDPKHWDGTARGWCVIGRLRYAA 209  
Db 163 FKPSGNVMPNP--GTTD-----FRDPKVIWDDQDNKVMVMAEGSKIG 205  
QY 210 FYTSPNLRDWTLLRRNFDPNHALGGTECPDLFEITADDGTRHVLAAASMDAYGIGLPMY 269  
Db 206 FRESNLDKWHYTSGE-FPEQA-GWCEPDLYMNRASDGTNRKWLJGASGKNGPWRKPNY 263  
QY 270 AYWTGTWDEQFHADLTLPQWLDGWDYAAVTPSIDA--PETKRLATAMNNKYYAAR 327  
Db 264 AYWTGSDCKEKFADQTEAQWLDYGDWYGGVTFEDSKSDTDPLEKRYALAWNNNDY-AN 322

QY 328 DVPTDASDNGNSIVRELRLARQCGGWTLLSTPVAALTNVYATTTLPDRTPDGSV 387  
Db 323 NTPT-MKNGFNGTDSVIRELKEQ-DGYTSLVSQIEALEQLTSTVSTDEIEDQVNGSKT 380  
QY 388 LPWNGRAYEIELDIAMDATNVGSGVSRSPDGRHNTNIGKY--GADLYVDGRGPDLAGYS 445  
Db 381 LSTTGTYQJDTDLSELSKAGVLRRESDDQRHLDVGLFAGGGYAYVNRATNPQDKS 440  
QY 446 LAPYSRAAAPIDPGARSVHLRLVLTQSVYEVFNAGHTVLSQQVHFAGDGLSYTDGG 505  
Db 441 -NTYVESKAPYDVNKKRKHVHLKILVDKTTIEVFVGDCKTFSNEVFPKPKDGITLYSDGG 499  
QY 506 PAHFTGIVR 515  
Db 500 TASFKNITVK 509

## RESULT 3

G97118

levanase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C:Accession: G97118

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
J. Dairy, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
Bacteriol. 183, 4823-4838, 2001A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: G97118

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 &lt;KUR&gt;

A:Cross-references: GB:AE001437; PIDN:AAK9738.1; PID:g15024743; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

C:Gene: CAC1773

C:Superfamily: Penicillium purporogenum inulinase

Query Match 31.0%; Score 880.5; DB 2; Length 514;  
Best Local Similarity 40.7%; Pred. No. 2e-57;  
Matches 200; Conservative 76; Mismatches 179; Indels 37; Gaps 14;

QY 40 RAVYHMTTPSGWLCQDPQPVTHGAYQLYLHS -DQNGPG-GWDHASTDGVAFTHHGT 97  
Db 45 REVYHFTVPMKNDPQRPVDFEGEYHYLYNKPNGNGTEWRQATSKDLYTWKDEGV 104  
QY 98 VMP--LRPDPFVMSGSAVVGTAAGAGAVVALATQPTDGVKRYEQOYLWSTDGGFT 155  
Db 105 CIPKYNENGDWGSFVVDQAQNTAGFGKNAIVATQPSASMDK-QEQLWSTDRGRT 163  
QY 156 FTALPD-PVIVNTDGRAATTPAEIENAEWFRDPKHWDGTARGWCVIGRLRYAAYTSP 214  
Db 164 FKPSNPIIPN-----PCTKD----FRDPKIIWDFKNNKVMVLAEGTKIGFVSY 211  
QY 215 NLRDWTLRRNFDPNHALGGTECPDLFEITADDGTRHVLAAASMDAYGIGLPMYAYWG 274  
Db 212 NLKNW--QHTGDFFTNIGIVECPDIFMNSDNGNTKWLGTSAKNGVSGEPNTYAYWG 269  
QY 275 TWDEQFHADLTLPQWLDGWDYAAVTPSIDAET--KRLAIAWNNKYYAARDVPTD 332  
Db 270 NYDGKFIADISTPKWLDYGFWDYAAVTFESENOSKRLKRYALAWNNNDY-ANNTP- 327  
QY 333 ASDGYNGNSIVRELRLARQCGGWTLLSTPVAALTNVYATTTLPDRTPDGSVLPWNG 392  
Db 328 IQNRFNGMSIVREITLSKQKNTYSLSKPIKRIENITTSIDQFKQISVKGLHLKVG 387  
QY 393 RAYEELTDAMDATNVGSGVSRSPDGRHNTNIGKY--GADLYVDGRGPDLAGYS LAP-- 448  
Db 388 KAYQDLDTDINWDKAKNVGIRLRESLDKKRHDVGIETEGKYSVNR-----AYTGNPK 441  
QY 449 ---YSRAAAPIDPGARSVHLRLVLTQSVYEVFNAGHTVLSQQVHFAGDGLSYTDGG 505  
Db 442 SKKYVESRAPFDLNNKRVHLRFVDKVSVEVFIDDKIYTSNEVFPKPKDGITLYS 501



|   |     |  |     |
|---|-----|--|-----|
| Db  | 258 | MOYFVGDFGTHFKNEPPNKNVLTWDGDRDFYAAVSWSDIPSDSRRLWLGWMSNQY-A      | 316 |
| Qy  | 327 | RDVPTDASDGYNGQNSIVRELRLARPGGWYTLTSTPVAALNIVYVATTTLTPORTVDGSA   | 386 |
| Db  | 317 | NDVPTSP---WRSATSI PRELKAKATEG-VRVQTPVKELETIRCTSKKKNLTISPAS     | 372 |
| Qy  | 387 | ---VLPWNGRAYEI--ELDIAMDATNVGISVGRSPDGRHRTNIGKYGADLYXDRGPSDL    | 441 |
| Db  | 373 | HNVLAGQSGDAYEINAEKFSVPGSAEEFGFKVRTGENQFTKGYDRNNAKLFVDRSEGN     | 432 |
| Qy  | 442 | AGYSIA-PYSRAAAPIDPGARSVHLRLIVDTQSVFVFNAGHTVTLISQVHFHAEAGDTGISL | 500 |
| Db  | 433 | DTFNPAPNTKETAPLKPNGVKLRFIDRSSRSEVFGNDGQVITDILPDRSKGLEL         | 492 |
| Qy  | 501 | YTDGGAHFTGIIVREI   | 517 |
| Db  | 493 | YAANG-----GVKXSL   | 504 |
| RESULT 6  |     |  |     |
| A36915  |     |  |     |
| fructanase - Bacteroides fragilis   |     |  |     |
| C:Species: Bacteroides fragilis   |     |  |     |
| C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000 |     |  |     |
| C:Accession: A36915   |     |  |     |
| R:Blatch, G.L.; Woods, D.R.   |     |  |     |
| J. Bacteriol. 175, 3058-3066, 1993  |     |  |     |
| A:Title: Molecular characterization of a fructanase produced by Bacteroides |     |  |     |
| A:Reference number: A36915; MUID:93259952; PMID:8491724                     |     |  |     |
| A:Contents: BF-1  |     |  |     |
| A:Accession: A36915   |     |  |     |
| A:Status: preliminary   |     |  |     |
| A:Molecule type: DNA  |     |  |     |
| A:Residues: 1-622 <BLA>   |     |  |     |
| A:Cross-references: GB:M83774; NID:gl43970; PIDN:AA822924.1; PID:gl43972    |     |  |     |
| A:Note: sequence extracted from NCBI backbone (NCBIN:131947, NCBI:P.131949) |     |  |     |
| C:Superfamily: beta-fructofuranosidase                                      |     |  |     |
| Query Match   |     |  |     |
| Best Local Similarity 30.2%; Pred. No. 2.8e-30;                             |     |  |     |
| Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps               |     |  |     |
| Qy  | 40  | RAVYHMTPPSGMLCDPQRPVTHGAYQLYXLSHDQNN--GPGGWDHASTTDCGVAFTHGT    | 97  |
| Db  | 133 | RPLYHHPTLYGWNDANGLVYKGEVHLFYQYNPYCSMGNNHWSKDLVHWEH---          | 189 |
| Qy  | 98  | VMPLRPDP-----VWGSAAVVGTTAGFAGAVVALATOPTDGVRYQEQYLWSTD          | 151 |
| Db  | 190 | ---LEPALARDTLGHIFSGSSVDDANTAGYGAGATVAFYTSASD--KNGQIQCMAYSTD    | 244 |
| Qy  | 152 | GGTFTTAL-PDPVIVTDCRAATTPAEIENAEWRDPKLIHWDTARGEVVCVIGRLRYAAF    | 210 |
| Db  | 245 | NGRTFTYKYNPLTVPDG-----LKDIFDPKVFVYAPQKVMVVSADKEMRF             | 293 |
| Qy  | 211 | YTPSLRDWTLRRNF-----YPNHALGIECPDLFEITADDCTRH--WVLAASMDAYGI      | 263 |
| Db  | 294 | YSSENLEKWTYMSGWEGYCVQPSQ-----FECPDWVLPDGNPHKKWALLIVNVN----     | 345 |
| Qy  | 264 | GLPMTY-----AYWTGTWDGEGFHADDL--TPQWLDGWDWYAAVTPWPSIDAPETKRLA    | 315 |
| Db  | 346 | --PCYFGGSATQYFIDGDFGEKFKVCDNKPETVWKLDMGKHATVCFESNTG---DRTIA    | 400 |
| Qy  | 316 | IAMNNHNYAARDVPTDASDGYNGQNSIVRELRLARQGGWYTLTSTPVAALTNVYVATT     | 375 |
| Db  | 401 | VPMMSNQY-ANIVPTQ---FRSNAALPRELSLTQDGDYI-MAAAPVEKTSLRKESR       | 455 |
| Qy  | 376 | TLPDRTVDGS---AVLPWNGRAYETELDIAWMTATNVGISVGRSPDGRHRTNIGKYGAD    | 431 |
| Db  | 456 | EIPAFEGVDAYHVDLSLLSDNKGAYELELAGSREIMGLKLFNEKGENVDIYISLPEKK     | 515 |
| Qy  | 432 | LYVDGRGSDLA--GYSLAPYSRAA-----APIDPGARSVHLRLV                   | 469 |

[illegible]

Db 604 NAEKWHLMVDQILOSDFKASGADVPR--LDYKDYAAVTDWNV--PNKRYOVGWS 660  
Qy 321 NKYAADVPDASDGYNGQNSIVRELRLARPGGWYILLSTPVAALNTYVTA-----T 374  
Db 661 NWAY--VRDLPITT---WRTAMSTVREMGLTR--VNGKRLRTAQPVTALSLRTOELIRKD 715  
Qy 375 TTLP--DRVVDGSAVLPWNGRAYEIELDIAMDATNVGIVSGRSPDGRHTNIG--KYGAD 431  
Db 716 TDIPVGETSLGKAA---QCTSLDISVLSPPASSPAGLKV--LDNGEQVTLIGYDSQAKQ 770  
Qy 432 LYVDRGPDLAGYSLAPYSRAAAPDPCAR--SVHLRIILVDTSQVEVFVFNAGHTVLSQOVH 490  
Db 771 LVDRTHSGVDFSPKFPARSTAPUSPSKGVHLRIILVDAHSVEVFAADGTPVITQTV 830  
Qy 491 FAEGDTGISLTXDGGPAHFTGIVVREIG 518  
Db 831 PQDATGVSLVAEGCTAHLGSLSLWHLG 858

RESULT 8  
JE0301  
inulinase (EC 3.2.1.7) - Aspergillus niger  
C:Species: Aspergillus niger  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 07-May-1999  
C:Accession: JE0301  
R:Ohta, K.; Akimoto, H.; Matsuda, S.; Toshimitsu, D.; Nakamura, T.  
BioSci. Biotechnol. Biochem. 62, 1731-1738, 1998  
A:Title: Molecular cloning and sequence analysis of two endoinulinase genes from Aspergillus niger  
A:Reference number: JE0301; MUID:95022189; PMID:9805373  
A:Accession: JE0301  
A:Molecule type: mRNA  
A:Residues: 1-516 <OHT>  
A:Cross-references: DDBJ:AB012771  
C:Superfamily: Penicillium purporogenum inulinase  
C:Keywords: glycosidase; hydrolase

Query Match 15.38; Score 434; DB 2; Length 516;  
Best Local Similarity 27.18; Pred. No. 2.le-24;  
Matches 144; Conservative 92; Mismatches 225; Indels 70; Gaps 19;

Qy 24 GGAVPPAARASAPGSLRAYVHMTPPSGWLCDDPQRVTHGAYOLYLLHSDQNGPGG--W 81  
Db 16 GLMPSQAQSN---DYRSYHFTPDQYWMNEPGLIKIGSTWHLEFQHNPTANWNGNCW 72  
Qy 82 DHASTDGVAFTHGTVMPLRPDPFVWNGSAGVGTANTAGFAGA---VVALATQPTDGV 138  
Db 73 GHATSTDLHWAHKAFTADENGVEAFTGTAYDPNNTSGLGDSANPPYLAFTGYTTS- 131  
Qy 139 RYQEQYLYWSTDGGFTTALPDVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEW 198  
Db 132 SQTQDRLAFSVDRNGATWTKFQGNPIITSQEA---PHDITGGLSREDPKVFFHRSQGNW 188  
Qy 199 VCVI--GRLRYAAFTSNLDRDWTLLRRNFDYPNHALGGI-----ECPDLFEITADDTG 249  
Db 189 IWLVAHGQDQLSFWTSADTINWTQW---DLKSTSLNGLSDDITGWEVPDMFELPV--EGT 245  
Qy 250 RH--WVL----AASMDAYGIGLPMYAYWTGTWGEQFADH--DLTPQWLDWGDWYAAV 301  
Db 246 EETTWVVMNTPAEGSPAGNGVLAI---TGSFDCGSFTADPVDASTMWDNGRDFDGL 301  
Qy 302 TWPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYLLS 361  
Db 302 SWANVPASDGRRIIAAVMNSV---GSPNPTTT---WKMLSPFRTLSL--KRVGTQQHFVQ 354  
Qy 362 TPVAALNTYVTTATLTPRTVDGSAVLPWNGRAYEIELDIADW--DTATNVGISVGRSPDG 419  
Db 355 QPITELDTISTLSLOTLANQITPGOTLSSIRGTALDVRVAFYDPDAGSVLSLAVRKGASE 414  
Qy 420 TRHNICKYGADLYVDRGPDLAGYSLAPYSRAAAPIDPGARSVH-----LR 466  
Db 415 QTVINYTQSNATLSVDRTESGDISY-----DPAAGGVHTAKLEEDGTGLVSIR 462  
Qy 467 ILVDTSQVEVFVFNAGHTVLSQOVHFAEGDGTGISLTDGGAHFTGIVVREI 517

Db 463 VLVDTSQVEVFEGGQGEAVISDLIFFSDSDGLEAVTGGNAVLSQVDSVRSV 513

RESULT 9  
JC7533  
inulinase (EC 3.2.1.7) - Penicillium sp. (strain TN-88)  
N:Alternate names: 2,1-beta-D-fructan fructanohydrolase  
C:Species: Penicillium sp. (strain TN-88)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7533  
R:Akimoto, H.; Kiyota, N.; Kushima, T.; Nakamura, T.; Ohta, K.  
Biosci. Biotechnol. Biochem. 64, 2328-2335, 2000  
A:Title: Molecular cloning and sequence analysis of an endoinulinase gene from Penicillium sp. (strain TN-88)  
A:Reference number: JC7533; MUID:21036892; PMID:11193399  
A:Accession: JC7533  
A:Molecule type: mRNA  
A:Residues: 1-515 <AKI>  
A:Cross-references: DDBJ:AB041337  
A:Experimental source: strain TN-88, wild type  
C:Comment: This enzyme, an acidic glycoprotein and an endo-acting inulinase as well, is  
asoe as the main products.  
C:Genetics:  
A:Gene: inuU  
C:Superfamily: Penicillium purporogenum inulinase  
C:Keywords: glycoprotein; glycosidase; hydrolase

Query Match 15.04; Score 425; DB 2; Length 515;  
Best Local Similarity 27.68; Pred. No. 9.9e-24;  
Matches 142; Conservative 92; Mismatches 232; Indels 48; Gaps 21;

Qy 31 ARASAGSLRAYVHMTPPSGWLCDDPQRVTHGAYOLYLLHSDQNN--GPGGWDHASTD 88  
Db 21 ARAVA--DDYRPAHFPCPAENWMNEPGLIQINSTWHLFYQADPAANWNGECWGHATSSD 79  
Qy 89 GVAFTHGTVMPLRDPDPFVWNGSAGVGTANTAGFAGA---VVALATQPTDGVRYQEQY 145  
Db 80 LLHWDHLVPAIVPENGIESFTGTSYSDNNTSGLTSTNPPYLAFTGYTES--NKTQDOR 138  
Qy 146 LYWSTDGGFTTALP--DPVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWCVI-- 202  
Db 139 LAYSTDLQGTWVKFAGNPII---GAAQEAQDISGGLSREDPKVFFHAPSCKWVMVLAH 194  
Qy 203 GRLRYAAFTSNLDRDWTLLRRNFD-----YPNHALGGIECDLFEITADDTGRH--WVL 254  
Db 195 GQODKLTEWTSLDKAKNWTWVSDLSSSQIEGFP--SITGWEVDMFQLPI--QCIKKTWVL 252  
Qy 255 ----AASMDAYGIGLPMYAYWTGTWGEQFADHDLTPQ--WLDWGDWYAAVTPSIDA 308  
Db 253 IFTPAOGSPPGNGV---VALTGSFDGETFVADPVPDFTLWLDYGRDFDGLSALWENVPA 308  
Qy 309 PETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARPGGWYLLSTPVAALT 368  
Db 309 SDGRRIIAAVMNSV---GSPNPTTT---WKMLSPFRTIAL--KQIGSKQYFLOQPVAE 361  
Qy 369 NYVTATTLTPRTVDGSAVL--PWNGRAYEIELDIAMDATNVGIVSGRSPDGRHTNIG 426  
Db 362 TIDGSLTSIQNTITPNTQLTSSIHCTSLDIRMAFVIDSGATLSLAVRKG--GSEQTIVIR 419  
Qy 427 KY--GADLYVDRGPDLAGYSLAPYSRAAAPIDPGARSVHLRIILVDTSQVEVFVFNAGHT 483  
Db 420 YFQSNSTLSVDRTAGSDISYDPAAGGVHTAQLAQNTELVLHIVALIDTCSVEVFGGEA 479  
Qy 484 VLSQOVHFAEGDGTGISLTDGGAHFTGIVVREI 517  
Db 480 VISDLIFFSNSDGLSLEVSGETAMLRSVNVSSV 513

RESULT 10  
C97080  
levanase/invertase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

Db 800 LAI 802

RESULT 11

JC5458

Inulinase (EC 3.2.1.7) precursor - Penicillium purporogenum  
N:Alternate names: 2,1-beta-D-fructan fructanohydrolase  
C:Species: Penicillium purporogenum  
C:Date: 17-Jun-1997 #sequence\_revision 17-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: JC5458; PC4326  
R:Onodera, S.; Murakami, T.; Ito, H.; Mori, H.; Matsui, H.; Honma, M.; Chiba  
Biosci. Biotechnol. Biochem. 60, 1780-1785, 1996  
A:Title: Molecular cloning and nucleotide sequence of cDNA and gene encoding  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-515/Product: endo-inulinase #status predicted <MAT>  
A:Reference number: JC5458; MUID:97141595; PMID:8987853  
A:Accession: JC5458  
A:Molecule type: DNA  
A:Residues: 1-515 <ON0>  
A:Cross-references: DDBJ; D84360; NID:g1805497; PIDN:BAAL2321.1; PID:g1805499  
A:Accession: PC4326  
A:Molecule type: protein  
A:Residues: 26-70;152-239;333-345;350-379 <ON02>  
C:Comment: This enzyme hydrolyzes inulin and fructo-oligo-saccharides, but  
C:Superfamily: Penicillium purporogenum inulinase  
C:Keywords: glycoprotein; glycosidase; hydrolase  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-515/Product: endo-inulinase #status predicted <MAT>  
F:108,109,132,246,424,489/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 13.7%; Score 389.5; DB 2; Length 515;  
Best Local Similarity 26.0%; Pred. No. 4.2e-21;  
Matches 133; Conservative 87; Mismatches 225; Indels 57; Gaps

Qy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLVYLSHDQNN--GPGGDHASTDTGVAFTHGT 97  
Db 29 RPTFHCPAENMNMENGLIKIDSTWHLFVQADPTANVWGNECHGATSSDULLHWHLPV 88  
Qy 98 VMLRPDPFPVWGSAYVGTANTAGFCAGA--VVALATQPTDGVKRYEQEYLWSTDGGF 154  
Db 89 AIPVENGIESFTGTSYDANNNTSSLGTSTNPFLAFTGYTSS--NGTDQRLAYSTDGTT 147  
Qy 155 TFTALP-DPVIYVNTDGRAATTAEIENAEWFRDPKIHWDTARGEWCVI--GRLLRYAIFY 211  
Db 148 TWLKFSGNPFI----SAALEAPHDVTGGLESRDPRKVFHPEPSKWMVYLAHGGGDKLTFW 203  
Qy 212 TSNLRDWTLLRNF-----DYPNHALGIECPDLFELTADDTGRH--WVL---AASMD 259  
Db 204 TSLDAKSWTMSDLSASQTEGPPS-SVYGWEVDPWFQPLI-OGTNETTWVLIPTPAOGSP 261  
Qy 260 AYGIGLPMTYAYWTGTWPDGEQFAD--DLTPQWLDWGDWYAAVTPSID:PETKRLAIA 317  
Db 262 AGNGV----VALTGSFDCGETFLANPVSSTLWLDYGRDFDGCAMSWENVPASDGLRIAA 317  
Qy 318 WMNWKYARDVPTDASDGYNGNSIVRELRLARQPGGWYLLSTPVAALNTNYTATTL 377  
Db 318 VMNSY---GSNPPTNT---WKGMLSPFRTLTLEKTSKQY-FLOQPIAELSTVDNALASI 370  
Qy 378 PDRTVGSVAVLPWNGRAYEIELDIADWTATNVGISVGRSPDGTHTNI--CKYGADLYVD 435  
Db 371 QNOTIAPKQTLLSSIHGSLDVRIFAISVDSGATLSLAVRKGGSEQTVIRYSQSNSTLSVD 430  
Qy 436 RGPSSDLAGYSLAPYSRAAAPIDPGARSVH-----LRLVDTQSVFVFNAGH 482  
Db 431 RTASGDISY-----DPAAGIHSQAARDNTLVLRLVLDVTCSEVFEVGQGE 478  
Qy 483 TVLSQQVHFAEGDTGSLSYTDGCGPAHFYGIIV 514  
Db 479 AVISDLIFFPSSNDGSLSEVIGTATLQSV 510

RESULT 12

A49206

exo-beta-D-fructosidase - Streptococcus mutans  
C:Species: Streptococcus mutans





A:Cross-references: EMBL:X07570

C:Genetics:

A:Gene: SUC1

A:Map position: 7R

C:Superfamily: beta-fructofuranosidase

C:Keywords: glycosidase; hydrolase

F:1-19/Domain: signal sequence status predicted <SIG>

F:20-532/Product: beta-fructofuranosidase 1 #status predicted <MAT>

Query Match 13.0%; Score 369.5; DB 2; Length 532;  
Best Local Similarity 25.1%; Pred. No. 1.3e-19;  
Matches 140; Conservative 84; Mismatches 221; Indels 113; Gaps 25;

QY 19 LALIFGAVPPAARASAPGSLRAYVHTPPSGWLCDPQR--PVTHGAYQLYVHSDQNN 76

Db 7 LFLLAGFAAKISAMTNETSDRLVHTPNKGMNDPGLWYDAKEGKWHLYFOYNPNDT 66

QY 77 G---PGWDHASTTDGVAFTTHG---TVMPLRPDPFVWGSAGVGTANTAGFGAGAV-- 127

Db 67 VWGLPLFWGHATSDD---LTHWQDEPVAIAPKRKDSGAYSGSMVIDYNNITSGFNDTIDP 123

QY 128 ---VALATQPTDGVRYQEOYLWSTDGGFTTAL-PDPVIVNTDGRAATTAEIENAE 182

Db 124 RQCVAIWYNTP---ESEEQYISYSLDGGYTFTEYQKNPVL-----ANST 167

QY 183 WFRDPKHWDTARGEWCWVIGRLR--YAAFTYSPNLRDWTLRNFDYPNHALGG--IECP 238

Db 168 QFRDPKRVFVWEPKSWIMTAARKSODYKIEIYSSDDLKSNKLESAP--ANEGLFGYQYCEP 225

QY 239 DLFEITA---DDGTRHWVLAASMDAYGIGLPMTYA---YWTGTWDEQFADDLTPQWLWD 293

Db 226 GLIEVPEQDPKSKHWVFMISNP---GAPAGGSFNQYFVGSFNGHFEAFDQNSRVDF 282

QY 294 GMDYAAVTPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRLAROP 353

Db 283 GRDYALQTFNTDPTYGSAAGLAWASNWEYSA-FVP---SNPWRSSMLVRPFSLNTE- 337

QY 354 GGYTLLSTPVAALTNVYATTTLPDRTVDGSAVLPMN-----GRAVEIELDIAM 403

Db 338 ---YQANPETELINLKA-----EPILNISAGPWSRFATNTTLTKANSYNVDLSNST 386

QY 404 DT-----ATNNGISVGRS-----PDGTRHTNIGKYGADLYVDRGPSOLA 442

Db 387 GTLEFLVYAVNTQTISKVSFADLSLWFKGLEDPPEYLRMGFEVSASFFLDRGNSKYK 446

QY 443 GYSLAPY--SRAAAPIDPGARSVHLRI-----LVDTSQSVFVFNAGHTVLSQQVHFAEGD 495

Db 447 FVKENPYFTNRMSVNNQPKSENDSLKYKYVGLDQNLILELYFNDGDVVYSTNTYFMTTCN 506

QY 496 -----TGIS--LYTD 503

Db 507 ALGSVNMTTGVDNLFYID 524

RESULT 15

S27373

beta-fructofuranosidase (EC 3.2.1.26) 4 precursor - yeast (Saccharomyces cerevisiae)

N:Alternate names: invertase; saccharase

C:Species: Saccharomyces cerevisiae

C:Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 22-Jun-1999

C:Accession: S27373; S25442

R:Hohmann, S.

submitted to the EMBL Data Library, November 1988

A:Reference number: S27372

A:Accession: S27373

A:Molecule type: DNA

A:Residues: 1-532 <OH>

A:Cross-References: EMBL:X07572; NID:94572; PIDN:CAA30459.1; PID:94573

R:Hohmann, S.; Gozalbo, D.

Mol. Gen. Genet. 211, 446-454, 1988

A:Title: Structural analysis of the 5' regions of yeast SUC genes revealed analogous pal

A:Reference number: S25439; MUID:88216256; PMID:2835632

A:Accession: S25442

A:Molecule type: DNA

A:Residues: 1-74 <HOH2>

A:Cross-references: EMBL:X07572

A:Note: the authors translated the codon ATT for residue 7 as Asn

C:Genetics:

A:Gene: SUC4

C:Superfamily: beta-fructofuranosidase

C:Keywords: glycosidase; hydrolase

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-532/Product: beta-fructofuranosidase 4 #status predicted <MAT>

Query Match 12.9%; Score 366.5; DB 2; Length 532;

Best Local Similarity 26.2%; Pred. No. 2.2e-19;

Matches 143; Conservative 80; Mismatches 231; Indels 91; Gaps 24;

QY 21 LIFGAVPPAARASAPGSLRAYVHTPPSGWLCDPQR--PVTHGAYQLYVHSDQNN-- 77

Db 9 LIAGFAAKISALMTNETSDRLVHTPNKGMNDPGLWYDAKEGKWHLYFOYNPNDTVW 68

QY 78 --PGWDHASTTDGVAFTTHG---TVMPLRPDPFVWGSAGVGTANTAGFGAGAV--- 127

Db 69 GLPLFWGHATSND---LTHWQDEPVAIAPKRNDGAYSGSMVIDHNTSEFFNDTVDPQ 125

QY 128 ---VALATQPTDGVRYQEOYLWSTDGGFTTAL-PDPVIVNTDGRAATTAEIENAE 184

Db 126 RCVAIWYNTP---ESEEQYISYSLDGGYTFTEYQKNPVL-----ANSTQ 169

QY 185 RPKIHWDTARGEWCWVIGRLR--YAAFTYSPNLRDWTLRNFDYPNHALGG--IECPDL 240

Db 170 RPKFVWEPKSWIMTAARKSODYKIEIYSSDDLKSNKLESAP--ANEGLFGYQYCEPGL 227

QY 241 FEITA---DDGTRHWVLAASMDAYGIGLPMTYA---YWTGTWDEQFADDLTPQWLWD 295

Db 228 IEVPTEDPDKSKHWVFMISNP---GAPAGGSFNQYFVGSFNGHFEAFDQNSRVDF 284

QY 296 DMYAAVTPSIDAPETKRLAIAMNNKYYAARDVPTDASDGYNGQNSIVRELRL----- 350

Db 285 DYALQTFNTDPTYGSAAGLAWASNWEYSA-FVPTNP---WRSSMSLVKXFSLNTEYOA 340

QY 351 RQPGWYTLSTPVAALTN-----YVATTTLPDRTVDGSAVLPMNGRAYEIELDIADWT 405

Db 341 NPETELINLKAEPILNISAGPWLHPASNTL-TKANSEVDLSNSTGTLEFLVAVNT 399

QY 406 ATNNGISV-----RSPDGRHTNIGKYGADLYVDRGPSOLAGYSLAPY---SRAA 453

Db 400 TQSVKSVFSDLSLWFKGLEDPPEYLRMGFEASASFFLDRGNSKYKFKENPYFTNRMS 459

QY 454 APIDPGARSVHLRI-----LVDTSQSVFVFNAGHTVLSQQVHFAEGD-----TGIS- 499

Db 460 VNNQPKSENDSLKYKYVGLDQNLILELYFNDGDVVYSTNTYFMTTGNALGSMVMTTGVDN 519

QY 500 --LYTD 503

Db 520 LFYID 524

Search completed: May 6, 2003, 12:35:49

Job time : 31.9442 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:23:52 ; Search time 10.4798 Seconds  
(without alignments)  
2054.066 Million cell updates/sec

Title: US-09-868-328B-4

Perfect score: 2837

Sequence: 1 MTPAIRRAVLQAGAGALA.....LYTDGGPAHFTGIVVREIGQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 556   | 19.6        | 677    | 1     | SACC_BACSU  |
| 2          | 382.5 | 13.5        | 1423   | 1     | FRUA_STRMU  |
| 3          | 379   | 13.4        | 550    | 1     | INV1_HANAN  |
| 4          | 369.5 | 13.0        | 532    | 1     | INV1_YEAST  |
| 5          | 366.5 | 12.9        | 532    | 1     | INV4_YEAST  |
| 6          | 365.5 | 12.9        | 555    | 1     | INV1_KLUMA  |
| 7          | 359.5 | 12.7        | 532    | 1     | INV2_YEAST  |
| 8          | 354   | 12.5        | 477    | 1     | CSCA_ECOLI  |
| 9          | 348   | 12.3        | 581    | 1     | INV1_SCHPO  |
| 10         | 330   | 11.6        | 476    | 1     | RAFD_ECOLI  |
| 11         | 320   | 11.3        | 533    | 1     | INV_DEBOC   |
| 12         | 312   | 11.0        | 590    | 1     | INVA_MAIZE  |
| 13         | 276   | 9.7         | 512    | 1     | INVA_ZYMO   |
| 14         | 274.5 | 9.7         | 636    | 1     | INVA_LYCES  |
| 15         | 260.5 | 9.2         | 640    | 1     | INV1_CAPAN  |
| 16         | 249   | 8.8         | 642    | 1     | INVA_VICFA  |
| 17         | 248   | 8.7         | 651    | 1     | INVA_PHAVU  |
| 18         | 246.5 | 8.7         | 465    | 1     | SCRB_KLEPN  |
| 19         | 246.5 | 8.7         | 555    | 1     | INV1_PEA    |
| 20         | 246   | 8.7         | 661    | 1     | INV1_DAUCA  |
| 21         | 243   | 8.6         | 670    | 1     | INV1_MAIZE  |
| 22         | 242   | 8.5         | 592    | 1     | INV1_DAUCA  |
| 23         | 242   | 8.5         | 649    | 1     | INVA_PHAAP  |
| 24         | 239.5 | 8.4         | 511    | 1     | SCRB_ZYMO   |
| 25         | 238.5 | 8.4         | 432    | 1     | BFRA_THEME  |
| 26         | 237.5 | 8.4         | 494    | 1     | SCRB_STAXY  |
| 27         | 232.5 | 8.2         | 484    | 1     | SCRB_VIBAL  |
| 28         | 222   | 7.8         | 454    | 1     | SCRB_STRMU  |
| 29         | 214   | 7.5         | 466    | 1     | SCRB_SALTY  |
| 30         | 213   | 7.5         | 473    | 1     | SCRB_LACLA  |
| 31         | 206   | 7.3         | 501    | 1     | SCRB_PEDPE  |
| 32         | 204.5 | 7.2         | 583    | 1     | INV3_DAUCA  |
| 33         | 197.5 | 7.0         | 480    | 1     | SCRB_BACSU  |

34 187.5 6.6 592 1 INV2\_DAUCA  
35 121 4.3 1120 1 STFR\_ECOLI  
36 117 4.1 451 1 Y483\_MYCTU  
37 117 4.1 3491 1 ERY1\_SACER  
38 116.5 4.1 984 1 EPA3\_RAT  
39 114.5 4.0 890 1 NIA2\_PHAVU  
40 114 4.0 1132 1 VHSJ\_LAMBD  
41 113.5 4.0 909 1 NIA\_PETHY  
42 113 4.0 296 1 ALYS\_BPDP1  
43 112 3.9 916 1 NIAL\_ORYSA  
44 111.5 3.9 752 1 HISA\_PROAC  
45 110.5 3.9 983 1 EPA3\_HUMAN

Q39692 daucus caro  
P76072 escherichia  
Q11149 mycobacteri  
Q03131 saccharopol  
O08680 rattus norv  
P39866 phaseolus v  
P03745 bacterioph  
P36859 petunia hyb  
O03979 bacterioph  
P16081 oryza sativ  
Q59634 propionibac  
P23320 homo sapien

#### ALIGNMENTS

RESULT 1

SACC\_BACSU  
ID SACC\_BACSU STANDARD; PRT; 677 AA.  
AC P05656;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan  
DE fructanohydrolase).  
GN SACC.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87286401; PubMed=3112519;  
RA Martin I., Debarbouille M., Ferrari E., Klier A., Rapoport G.;  
RT "Characterization of the levanase gene of Bacillus subtilis which  
RT shows homology to yeast invertase.";  
RL Mol. Gen. Genet. 208:177-184(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=168;  
RA MEDLINE=88067786; PubMed=3120151;  
RA Schoergendorfer K., Schwab H., Lafferty R.M.;  
RT "Nucleotide sequence of a cloned 2.5 kb PstI-EcoRI Bacillus subtilis  
RL DNA fragment coding for levanase.";  
RN Nucleic Acids Res. 15:9606-9606(1987).  
RP SEQUENCE FROM N.A.  
RX STRAIN=168;  
RA MEDLINE=97286548; PubMed=9141695;  
RA Parro V., San Roman M., Galindo I., Purnelle B., Bolotin A.,  
RA Sorokin A., Mellado R.P.;  
RT "A 2391 bp region of the Bacillus subtilis genome comprising genes  
RT located upstream and downstream of the lev operon.";  
RL Microbiology 143:1321-1326(1997).  
RN [4]  
RP SEQUENCE OF 1-10 FROM N.A.  
RX MEDLINE=90355183; PubMed=2117666;  
RA Martin-Verstraete I., Debarbouille M., Klier A., Rapoport G.;  
RT "Levanase operon of Bacillus subtilis includes a fructose-specific  
RT phosphotransferase system regulating the expression of the operon.";  
RL J. Mol. Biol. 214:657-671(1990).  
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 2,6-beta-D-  
CC fructofuranosidic linkages in 2,6-beta-D-fructans (Levans)  
CC containing more than 3 fructose units.  
CC -1- INDUCTION: IN CONTRAST WITH LEVANSUCRASE AND SUCRASE, LEVANASE  
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE; NO INDUCER IS KNOWN FOR  
CC LEVANASE.  
CC -1- MISCELLANEOUS: LEVANASE CANNOT BE DETECTED IN THE WILD-TYPE  
CC B. SUBTILIS BUT IS MOSTLY SECRETED INTO THE CULTURE MEDIUM BY SACL  
CC MUTANTS, ESPECIALLY AT THE END OF THE EXPONENTIAL GROWTH PHASE.  
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X05649; CAA29137.1; ALT\_INIT.  
DR EMBL; Y00485; CAA68542.1; -  
DR EMBL; X92868; CAA63465.1; -  
DR EMBL; X56098; CAA39581.1; -  
DR EMBL; Z99117; CAB14645.1; -  
DR PIR; S06353; S06353.  
DR PIR; A27286; A27286.  
DR Subtilist; BG10320; sacC.  
DR InterPro; IPR001362; GH\_32.  
DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Hydrolase; Glycosidase; Signal; Complete proteome.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 677 LEVANASE.  
FT ACT\_SITE 49 49 BY SIMILARITY.  
FT CONFLICT 658 658 O -> L (IN REF. 2).  
SQ SEQUENCE 677 AA; 75951 MW; 80FD680A5EE7F525 CRC64;  
  
Query Match 19.6%; Score 556; DB 1; Length 677;  
Best Local Similarity 32.2%; Pred. No. 9.3e-33;  
Matches 160; Conservative 74; Mismatches 217; Indels 46; Gaps 18;  
  
QY 40 RAVYHMTPPSGWLCDDPQRPVTHHGYQLVYHSDQNN--GPGGWDHASTTDCGVAFTHCT 97  
DB 35 RQYHFTPPANMNDPVGMYAGETHLYFYHYPYGLQWPHWGHAVSKDLVTWCH--L 92  
QY 98 VMPLRPD--FPVWSGSAVGTANTAGFAG---AVVALATQPTDGVKRYQEYLYWSTDG 152  
DB 93 PVALYPDEKGTIFSGSAVDKNTSGFTGKEKPLVAITYQREG---HOVSIAVNDK 149  
QY 153 GTFTALP--DPVIVNFDGRAATTPAEIENAEWFRDPKIHWDTARGEWVCVIGRLRYAAPY 211  
DB 150 GRTWTKYAGNPVTPN-----PGKDD---FRDPKVPWEKEKKWVLAAGDRILIY 197  
QY 212 TSPNLRDWTLRNFDYPNHALGI--ECPDLFEITAD--DGRHWVLAASMDAYGICLPWT 268  
DB 198 TSKNLKQWITYASEFGQDQSGHGWECPDLPFLVDNPNQKKNQVSVGNCAVSGSG 257  
QY 269 YAYWTGTWDEGFHADDTLPQ--WLDWGDWYAATWPSIDAPETKRLAIANWNNKYYAA 326  
DB 258 MQYFVCDFTGHEKNENPNKVLWTDYGRDFAAYVSWSDIPSTDRLRLGLWNSNWOY-A 316  
QY 327 RDVPTDASDGYNGQNSIVRELRLAROPGWYTLTSTPVAALTYNYVATTTLPDRTVDGSA 386  
DB 317 NDVPTSP---WRSATSIPELKLKAFTEG--VRVQVTPVKELETIRGTSKKKNLITSPAS 372  
QY 387 ---VLPWNGRAVEI--ELDIANDTATNVCISVGRSPDGTTRHNTIGKYADLYVDGRPSDL 441  
DB 373 HNVLAGQSDAYEINAEFVPGSAAEFGFKVYRTGENQTKYGYDRRNNAKLEFVDRSEGN 432  
QY 442 AGYSILA-PYSRAAPIDPGARSVHLRLVDTQSVFVNAHTVLSQOVHFAEGDGTISL 500  
DB 433 DTFNPAFTNGKETAPLKPNGVKVRLRIFVDRSSVEVFGNDGKQVITDILPDRSSKGLEL 492  
QY 501 YTDGGPAHTGIVVREI 517  
DB 493 YAANG-----GVKVKSL 504  
  
RESULT 2  
FRUA\_STRMU  
ID FRUA\_STRMU STANDARD; PRT; 1423 AA.  
AC Q03174;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Fructan beta-fructosidase precursor (EC 3.2.1.80) (Exo-beta-D-  
DE fructosidase) (Fructanase).  
GN FRUA.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS-5;  
RX MEDLINE=93014178; PubMed=1398976;  
RA Burne R.A., Penders J.E.C.;  
RT "Characterization of the Streptococcus mutans GS-5 frua gene encoding  
RL exo-beta-D-fructosidase.";  
RL Infect. Immun. 60:4621-4632(1992).  
CC -!- FUNCTION: THIS PROTEIN IS A FRUCTANASE ENZYME WHICH DEGRADES  
CC LEVANS AND INULINS TO FRUCTOSE AND ALSO CLEAVES SUCROSE INTO  
CC GLUCOSE AND FRUCTOSE AND CAN THEREFORE FUNCTION AS AN  
CC EXTRACELLULAR INVERTASE.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal-non-reducing 2,1- and  
CC 2,6-linked beta-D-fructofuranose residues in fructans.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan  
CC by an amide bond (Potential).  
CC -!- INDUCTION: BY SUCROSE, FRUCTAN SUBSTRATES AND FRUCTOSE.  
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U78296; AAA26889.1; -  
DR InterPro; IPR003343; Big\_2.  
DR InterPro; IPR001362; GH\_32.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
DR Pfam; PF02368; Big\_2; 1.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; FALSE\_NEG.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; FALSE\_NEG.  
KW Hydrolase; Glycosidase; Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1 39 POTENTIAL.  
FT CHAIN 40 1391 FRUCTAN BETA-FRUCTOSIDASE.  
FT PROPEP 1392 1423 REMOVED BY SORTASE (POTENTIAL).  
FT ACT\_SITE 458 458 BY SIMILARITY.  
FT DOMAIN 867 871 INVOLVED IN BINDING OF SUGARS WITH  
FT BETA-(2,6) LINKAGES OR BINDING OF  
FT MOLECULAR WEIGHT FRUCTANS (BY  
FT SIMILARITY).  
FT SITE 1388 1392 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD\_RES 1391 1391 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 1423 AA; 158668 MW; 8E574715F4E72A8A CRC64;  
  
Query Match 13.5%; Score 382.5; DB 1; Length 1423;  
Best Local Similarity 28.2%; Pred. No. 7.6e-20;  
Matches 148; Conservative 61; Mismatches 193; Indels 123; Gaps 27;  
  
QY 40 RAVYHMTPPSGWLCDDPQRPVTHHGYQLVY-LHSDNNGPGGWDHASTTDCGVAFTHCTV 98  
DB 444 RDQYHYSVKDGNWDPNGLVYNGVYHLPFHQFYDDTKGPMHWAHATSTDLI-----HWKE 499  
QY 99 MPLR--PDFP---VWMSGVAVGTANTAGF---GAGAVVALATQPTDGVKRYQEYLYWSTD 151  
DB 500 EPIAFYPSDNGYMFSGCVVVDHNSGGLKTKAGGLVAITANGNG-----QRMELAYSED 555  
QY 152 GGFTFTALPDPVIVNTDGRAATTPAEIENAEW-----FRDPKI--HWDTARGEWVCV 201  
DB 556 ECKTWQKY--DRIV-----ADWSNDPLQNDQFDPKPVFHN---NQWFW 595  
QY 202 I--GRURYAAFTYSPNLRDWTLRNFDYPN--HALGIECPDLFEITADGTRHWVLAAS 258

```
Db 596 LAGGFLR---YSSNNLKDWKVEST--YPLDHT-----ECPDMYPIVANDGVKLKVLRSRG 646
Qy 259 DAYGIGLPMYATYGTWDEGEFHADDL--TPWLDGWDYAAVW-----PS 305
Db 647 REYKVG---DFKQVDGKW---TFIADDAKKQDQVNFEGKDYAAWTVYVDFGTETRT 700
Qy 306 IDAPETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLTSTPVA 365
Db 701 I-----PKLTEVNMWMTWEDYCNLADVTVGQDFNGTFNLNLDGLINE-NGQYILTQTPVK 755
Qy 366 ALTNVVTATTL--PDRVDSAVL--PWNGRAYEIELDIAMDATNTNNGISVGRSPDGT 421
Db 756 AYDSURDYNTALHFQKDVTVDAANTLLKDFKGSYIEVSHFRPEDETT-----802
Qy 422 HTNICKYGADLVVDRGSDLAGYSLAP-----YSRAAPIDP-----GARS 462
Db 803 -----KVGFNLRVNGQATKVYIDLOTLSIDRSQSGTILSAFAKVNQHVTRNADGS 857
Qy 463 VHLRLVPTQSVVEFVFNAGHTVLSQQVHFAEGDTGISLYTDGSPA 507
Db 858 IDLHIYVDRASVEFVKNTVAGANQIFPNPAVCAIIVEGKA 902

RESULT 3
INV1_HANAN
ID INV1_HANAN STANDARD; PRT; 550 AA.
AC P40912;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)
DE (Saccharase).
GN INV1.
OS Hansenula anomala (Yeast) (Candida pelliculosa).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4927;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96163052; PubMed=8595669;
RA Perez J.A.; Rodriguez J.; Rodriguez L.; Ruiz T.;
RT "Cloning and sequence analysis of the invertase gene INV 1 from the
RT yeast Pichia anomala";
RL Curr. Genet. 29:234-240(1996).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X80640; CAA56684.1; -
CC PIR; S48812; S48812.
CC IncerPro; IPR001362; GH_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
CC Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 550 INVERTASE.
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
```

```
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
---FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 63188 MW; 192BCD722455E121 CRC64;

Query Match 13.4%; Score 379; DB 1; Length 550;
Best Local Similarity 24.6%; Pred. No. 4.3e-20;
Matches 139; Conservative 87; Mismatches 212; Indels 128; Gaps 27;

Qy 39 LRAVYHMTPPSQWLCDPQRPVTHGAY-----QLYVLHSDQNNG-----PGGWDHAST 86
Db 27 LRPOHLHPDQGMNDP-----NMFYDRDKLKHVVFQHPDPKKSIIWATPVTHGHSIS 80
Qy 87 TDGVAFTHHGTVM-PLRDPFVWMSGVAVGTANTAGFGAGA-----VVALATQPTDGV 139
Db 81 KDLLTWYHGNALPENDEDEGIFSGSVVDRNRTSGFFNDSTDPEQRIVAIYTNA----136
Qy 140 KYOEQYLWSTDGGTFTALPDPPVIVNTDGRAATPAEIEAENAEWRDPKIHWDTARGEW 199
Db 137 QLQTEIAIYSLDKYSFKYDQNPVINV-----NSSQQRDPKVLWHDESNOWI 184
Qy 200 CVIGRLR--YAAFYTSPLNRDMWTLRRNFDYPNHALGGIECPDLFEITA-----DDGTRHW 252
Db 185 MYVAKTQEFKVOIYGSPDLKKWDLKSNFTSNGYLGFEQECPLFLPIENPLNDIVTSKW 244
Qy 253 VLAASMDAYGIGLPM---TYAYWTGTWDEGEFHADDLTPQWLWDGMDWYAAVTHPSIDAP 309
Db 245 VLLALNP---GSLPGGSINEYFIEDGDTTFHPDGGATREMDICKDFYAFQSEFNTPE-P 300
Qy 310 ETKRLAIAMNNWYAAARDVPTDASDGYNGQNSIVRELRLAR---OPGGW-YTLTSTPVA 365
Db 301 EDGALGLAWASNMQY-ANTVPT---ENWRSSMSLVNRYTLKYVDVNPENYGLTLTKQPV- 355
Qy 366 ALTNVVTATTL-----PDRVDSAVLPPWNGRAYE-----LELDIAMD 405
Db 356 ----YDTETRLNETLKTLETINEYEVNDLKLKSSFVATDENTERNATGVEFDLKP---409
Qy 406 ATNVGISVGRSPDGTTRHTNIGKY-----GADLYVDRGP-----438
Db 410 -TQTDLKMGS---NMTTQFGLYIHSQTVKGSQETLQLVDFTLSTTWYIDRTTQHSFORN 465
Qy 439 SLAGYSLAPYSRAAPIDPGARSVH-LRLVDTQSVEFVFNAGHTVLSQQVHFAEG--D 495
Db 466 SPVFTERISTVYVEKIDTTDQG--NVYTLGVVDRNILELYFNDGSIAMTNTFFREGKIP 523
Qy 496 TGISLYTDGGAHAFT--GIVVREIGO 519
Db 524 TSFEVVCDSKSFIDELSVRELAR 549

RESULT 4
INV1_YEAST
ID INV1_YEAST STANDARD; PRT; 532 AA.
AC P10594;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Invertase 1 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 1)
DE (Saccharase).
GN SUC1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-73 FROM N.A.
RX MEDLINE=88216256; PubMed=2835632;
RA Hohmann S.; Gozalbo D.;
RT "Structural analysis of the 5' regions of yeast SUC genes revealed
RT analogous palindromes in SUC, MAL and GAL.";
RL Mol. Gen. Genet. 211:446-454(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Hohmann S.;
```

|              |   |  |                                     |
|--------------|---|--|-------------------------------------|
| QY           | 443   | GYSLAYD--SRAAPIDPGARSVHLRI-----LVDTQSVEVFVNAGHTVLVSQQVHFAEGD | 495                                 |
|              |   | :  :   | :  :    :  :                        |
| Db           | 447   | FVKENPYFTNRMSVNNQPKSENDLSVYKVGILLDNILELYFENDGVSVSTNTFYEMTTGN | 506                                 |
| QY           | 496   | -----TGIS--LYTD  | 503                                 |
|              |   | :  | :                                   |
| Db           | 507   | ALGSVNMVTGVDNLFYID   | 524                                 |
| <br>RESULT 5 |   |  |                                     |
| ID           | INV4_YEAST  | STANDARD;  | PRT; 532 AA.                        |
| AC           | PI0596;   |  |                                     |
| DT           | 01-JUL-1989 (Rel. 11, Created)  |  |                                     |
| DT           | 01-JUL-1989 (Rel. 11, Last sequence update)                                   |  |                                     |
| DT           | 01-MAR-1992 (Rel. 21, Last annotation update)                                 |  |                                     |
| DE           | Invertase 4 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 4) (Saccharase). |  |                                     |
| DE           | SUC4.   |  |                                     |
| GN           | Saccharomyces cerevisiae (Baker's yeast).                                     |  |                                     |
| OS           | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;              |  |                                     |
| OC           | Saccharomycetales; Saccharomycetaceae; Saccharomycetes.                       |  |                                     |
| NCBI_TaxID   | 4932;   |  |                                     |
| RN           | [1]   |  |                                     |
| RP           | SEQUENCE OF 1-73 FROM N.A.  |  |                                     |
| RX           | MEDLINE=88216256; PubMed=2835632;   |  |                                     |
| RA           | Hohmann S.; Gozalbo D.;   |  |                                     |
| RT           | "Structural analysis of the 5' regions of yeast suc genes revealed            |  |                                     |
| RT           | analogous palindromes in suc, mal and gal.";                                  |  |                                     |
| RL           | Mol. Gen. Genet. 211:446-454(1988).   |  |                                     |
| RN           | [2]   |  |                                     |
| RP           | SEQUENCE FROM N.A.  |  |                                     |
| RA           | Hohmann S.;   |  |                                     |
| RL           | Submitted (NOV-1988) to the EMBL/GenBank/DBJ databases.                       |  |                                     |
| CC           | -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-           |  |                                     |
| CC           | fructofuranoside residues in beta-D-fructofuranosides.                        |  |                                     |
| CC           | -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.                  |  |                                     |
| CC           | ----  |  |                                     |
| CC           | This SWISS-PROT entry is copyright. It is produced through a collaboration    |  |                                     |
| CC           | between the Swiss Institute of Bioinformatics and the EMBL outstation -       |  |                                     |
| CC           | the European Bioinformatics Institute. There are no restrictions on its       |  |                                     |
| CC           | use by non-profit institutions as long as its content is in no way            |  |                                     |
| CC           | modified and this statement is not removed. Usage by and for commercial       |  |                                     |
| CC           | entities requires a license agreement (See http://www.isb-sib.ch/announce/    |  |                                     |
| CC           | or send an email to license@isb-sib.ch).                                      |  |                                     |
| CC           | ----  |  |                                     |
| DR           | EMBL; X07572; CAA30459.1;   |  |                                     |
| DR           | PIR; S27373; S27373.  |  |                                     |
| DR           | SGD; L0002146; SUC4.  |  |                                     |
| DR           | InterPro; IPR001362; GH_32.   |  |                                     |
| DR           | Prfam; PF00251; Glyco_hydro_32; 1.  |  |                                     |
| DR           | PROSITE; P500609; GLYCOSYL_HYDROL_F32; 1.                                     |  |                                     |
| KW           | Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.               |  |                                     |
| FT           | SIGNAL  | 1 19   |                                     |
| FT           | CHAIN   | 20 532   | INVERTASE 4.                        |
| FT           | ACT_SITE  | 42 42  | BY SIMILARITY.                      |
| FT           | CARBOHYD  | 23 23  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 64 64  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 97 97  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 111 111  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 112 112  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 118 118  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 165 165  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 266 266  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 275 275  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 356 356  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 369 369  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 384 384  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 398 398  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | CARBOHYD  | 512 512  | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT           | SEQUENCE  | 532 AA; 60575 MW; A5GEA2E3A49125EB CRC64;                    |                                     |

12.98; Score 366.5; DB 1; Length 532;

[illegible]



Db 150 ---QAQHSISLGGYTFEHTYDNAVLDI-----NSSNFRDPKVFVHEGENCE 194  
QY 196 -GEWVCVIGRLRY--AAFYTSPLNRDWTLLRRNFDYDPNHALGGIECPDLFEITAD---DGT 249  
Db 195 DGRWIMAVAESQVFSVLFSYSSPNLKNWLTSTHGHGWTGTYECPGLVKVPYDSSVVDSS 254  
QY 250 RH-----WVLAASMDAYG-TGLPMTYAYWTGTWDGEOFHADDLTLPQWLDNGWDMWAA 300  
Db 255 NSSDSKPSAWLVSINPGGPGGSVT-QYFVGDFNGHTFTPTDGTQRELDMGKDYAL 313  
QY 301 VTPSIDAPETKRL-ATAMNNKRYAARDVPTDASDGYNGQNSIVRELRL---AROPGGW 356  
Db 314 QTF--FWTPEKDYVGIYASNNQY-AQAQPTDP---WRSSMLVROFTLUKDFSTPNNSA 367  
QY 357 YTLT-STPVAALNTYVATTTLTDPRTVDGSAVLPMWNGRAVEIETDIAWDATNNGISVGR 415  
Db 368 DVLNSQPV---LNYDALRNGTYTSTNTVTSSENGK---KIKLD----- 407  
QY 416 SPQTRHTNTKYGADLYDRGSDIAGYSLAPYSRAAAPIDPGARSVHLRLVDTQSVSE 475  
Db 408 NPSSGSEFHE-----YVFNPGSPDKSNVFAADLSLYFKGNDD--NEYRLRGYETNGGA 459  
QY 476 VEYNAGHT 483  
Db 460 FFLDRGHT 467  
RESULT 7  
INV2\_YEAST STANDARD; PRT: 532 AA.  
AC P00724;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Invertase 2 precursor (EC 3.2.1.26) (Beta-fructofuranosidase 2)  
DE (Saccharase)  
GN SUC2 OR YII162W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83168934; PubMed=6300785;  
RA Taussig R., Carlson M.;  
RT "Nucleotide sequence of the yeast SUC2 gene for invertase.";  
RL Nucleic Acids Res. 11:1943-1954(1983);  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
RA Walsh S.V., Whitehead S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-30 FROM N.A.  
RX MEDLINE=85137467; PubMed=6396505;  
RA Sarokin L., Carlson M.;  
RT "Upstream region required for regulated expression of the glucose-  
repressible SUC2 gene of Saccharomyces cerevisiae.";  
RL Mol. Cell. Biol. 4:2750-2757(1984).  
RN [4]  
RP SEQUENCE OF 1-21 FROM N.A.  
RX MEDLINE=87064535; PubMed=3537718;  
RA Kaiser C.A., Botstein D.;  
RT "Secretion-defective mutations in the signal sequence for  
Saccharomyces cerevisiae invertase.";  
RL Mol. Cell. Biol. 6:2382-2391(1986).  
RN [5]  
RP SEQUENCE OF 21-532, AND CARBOHYDRATE-LINKAGE SITES.

RX MEDLINE=88213364; PubMed=3284881;  
RA Reddy V.A., Johnson R.S., Biemann K., Williams R.S., Ziegler F.D.,  
RA Trimble R.B., Maley F.;  
RT "Characterization of the glycosylation sites in yeast external  
invertase. I. N-linked oligosaccharide content of the individual  
sequons.";  
RL J. Biol. Chem. 263:6978-6985(1988).  
RN [6]  
RP ACTIVE SITE ASP-42.  
RX MEDLINE=90293006; PubMed=2113524;  
RA Reddy V.A., Maley F.;  
RT "Identification of an active-site residue in yeast invertase by  
affinity labeling and site-directed mutagenesis.";  
RL J. Biol. Chem. 265:10817-10820(1990).  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
fructofuranoside residues in beta-D-fructofuranosides.  
CC -I- PTM: THE ENZYME EXISTS IN TWO STATES: A GLYCOSYLATED EXTERNAL  
FORM, OR A NON GLYCOSYLATED INTERNAL FORM.  
CC -I- MISCELLANEOUS: THE SEQUENCE OF THE SECRETED FORM OF THE ENZYME IS  
SHOWN.  
CC -I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 246921; CAA87030.1; -  
DR EMBL; V01311; CAA24618.1; -  
DR EMBL; K03294; AAA35127.1; -  
DR EMBL; M13627; AAA35129.1; -  
DR PIR; A00899; IFBY  
DR PIR; A27748; A27748.  
DR GlycosuiteDB; P00724; -  
DR COMPLEYEAST-2DPAGE; P00724; -  
DR SGD; S0001424; SUC2.  
DR InterPro; IPR001362; GH\_32.  
DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 532  
FT ACT\_SITE 21 532  
FT ACT\_SITE 42 42  
FT CARBOHYD 23 23  
FT CARBOHYD 64 64  
FT CARBOHYD 97 97  
FT CARBOHYD 111 111  
FT CARBOHYD 118 118  
FT CARBOHYD 165 165  
FT CARBOHYD 266 266  
FT CARBOHYD 275 275  
FT CARBOHYD 356 356  
FT CARBOHYD 369 369  
FT CARBOHYD 384 384  
FT CARBOHYD 398 398  
FT CARBOHYD 512 512  
FT MUTAGEN 42 42  
FT CONFLICT 409 409  
SQ SEQUENCE 532 AA; 60639 MW; 7D8AB33E6772B775 CRC64;  
Query Match 12.7%; Score 359.5; DB 1; Length 532;  
Best Local Similarity 26.0%; Pred No. 1,le-18;  
Matches 143; Conservative 81; Mismatches 228; Indels 99; Gaps 25;  
QY 19 LALIFGAVPPAARASAPGSLRAVYHMTPPSGWLCDDP-----QRPVTHGAVQLYLLHS 72  
Db 7 LFLAGFAAKISASMTNETSDRPLVHFTPNKGMNDPGLWYDEKDAKWH---LYFQYN 62  
QY 73 DQN---NGPGGWDHASTTGDGVAFTHHG-TVMPLRPFVWGSAAVVGTAAGFGAGAV- 127

```
Db 63 PNDVTGTPLEFCHGATSDDLTNWEDQIAIAPKRNDSGAFSGSMVVDVNTSGFFNDTRD 122
Qy 128 -----VALATOPTDGVRYQEOYLXWSTDDGGFTTAL-PDPVIVNTDGRAATTPAEIEA 181
Db 123 PRQRCVATWYNT-EEEOYISYLDGGYTFTEYOKNPVLA-----ANS 166
Qy 182 ENFRDPKHMTDARGWCVIGRLR--YAAFTSPNLRDWTLLRRNFDYPNHALGG--TEC 237
Db 167 TOFRDPKVFWEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESFAF--ANEGLGQYEC 224
Qy 238 PDLFEITA--DDGTRHWLAASMDAYGIGLPMTYA--YWTGTWDEGFHADDTLPQWLD 292
Db 225 PELIEVTEQDPSKSYWMEFISNP--GAPAGGSFNOYFVGSFNGTFHEAFDQNSRVVD 281
Qy 293 WGDWYAAATWPSIDAPETKRLAIAMNNWKYAADVPTDASDGYNGQNSIVRELRL--- 349
Db 262 FGKDYALQTFNTDPTVGSALGIANASWEISA-FVTNP---WRSSLSVKRKSLENT 337
Qy 350 --AROPGGWYTLTSPVAALN-----YVTATTLTPORT--VDGSVLPWNGRAYEIEL 399
Db 338 YOANPETELINLKARPILNISNAGPWSRFATNTTLTKANSYNVD-----LSNSTGTLEFEL 393
Qy 400 DIAWTATNVCISVG-----RSPDGRHTNIGKYGADLYVDGRPSDLAGYSLAPY 449
Db 394 YVAVNTTQTSKSVFADLSLWFKGLEDPPEYLRMGFEVSASFLLDRGNSKYKFKENPY 453
Qy 450 --SRAAAPIDPGARSVHLRI-----LVDTQSVSEVFNAGHTVLSQQVHFAEGD----- 495
Db 454 FTRNSVNNQPKSENDSYKYVGYLLDQNLLELFNDGDVVVSTWYFTWYTGALGSVNM 513
Qy 496 -TGIS--LYTD 503
Db 514 TTGVNDLEVID 524

RESULT 8
CSCA_ECOLI
ID CSCA_ECOLI STANDARD; PRT; 477 AA.
AC P40714;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sucrose-6-phosphate hydrolase (EC 3.2.1.26) (Sucrase) (Invertase).
GN CSCA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC3132;
RA Bockmann J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS ENZYME ENABLES THE BACTERIA TO METABOLIZE SUCROSE
CC AS SOLE CARBON SOURCE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -!- PATHWAY: Sucrose metabolism
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X81461; CAA57219.1; -
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
```

```
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism.
FT ACT_SITE 39 BY SIMILARITY.
SQ SEQUENCE 477 AA; 54363 MW; 48E6622AEA380EBD CRC64;

Query Match 12.5%; Score 354; DB 1; Length 477;
Best Local Similarity 27.2%; Pred. No. 2.3e-18;
Matches 135; Conservative 67; Mismatches 206; Indels 88; Gaps 23;

Qy 43 YHMTPPSGWGLDQPPQVTHTGAYQLYLIHSDQNN--GPGGMDHASTTQGVAFTHHGTVM 100
Db 28 FHLAPPAGWMDPNGLIWFNDRIYHAFYOHHPMSEHWGPMHGHATSDDMIHWQHE--PIA 85
Qy 101 LRPDPP-----VWSSGSAVYVNTANTAGFAGAVVALATOPTDGVRRYQEOYLXWSTDDGG 155
Db 86 LAPGDENDKDCGFCSSAVDDNGVLSLIYTGHWLDGAGNDDAIR--EVOCLATSRD--- 140
Qy 156 FTALDPPVIVNTDGRAATTPAEIEAENAFRDPKIHWDTARGEWVCV-----IGRLRYAAY 211
Db 141 -----IHFEKQGVILTPP---EGIMHFRDPKV--WREADTWMMVVGAKDPGNTGQILLY 189
Qy 212 TSPNLRDWTLLRNFDYPNHALGG---TECPDLREITADDGTRHWVLA--SMDAYGIGL 265
Db 190 RGSSUREWTFDRVL---AHADAGESYMWECDDFFSL-----GDQHYLMFSPOGMNAEG--- 239
Qy 266 PMTYAYWT-----GTWDGEQFHADDLTPQWLDWGDWYAAVTPSIDAPETKRLAIA 317
Db 240 ---YSYRNRFQSGVIVPGWMSPCRLFAQSGHGTDLNHDGFYAP---QSFVAKDGRRIYIG 293
Qy 318 WNNWKYAADVPTDASDGYNGQNSIVRELRLAROPGGWYTLTSPVAALNRYVATTTTL 377
Db 294 WMDWE---SPMPS--KREGWAGCMTLAREL---SESNG--KILQRPVHEAESLROQHOSI 344
Qy 378 PDRTVDGSAVLPWNGRAYEIELDIATW--DTATNVGISVGRSPDGRHTNIGKYGADLYVD 435
Db 345 SPRTISNKYVLOENAAQAVEIQWALKNSDAEHYGLQIGA-----GMRLYID 391
Qy 436 RGPDSLAGYSLAPYSR-----AAAPIDPGARSVHLRIILVDTSQSVFVFNAGHTVLSQQVHF 491
Db 392 NQSERLVLRWYYPHENLDGYSRISPL-POGDMALRIFDITSSVEVFINDGEAVMSSRIYP 450
Qy 492 AEGDTGISLYTDGGPA 507
Db 451 QPEERELSLYASHGVA 466

RESULT 9
INV1_SCHPO
ID INV1_SCHPO STANDARD; PRT; 581 AA.
AC O59852; P78891;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Invertase precursor (EC 3.2.1.26) (Beta-fructofuranosidase)
DE (Saccharase).
GN INV1 OR SPCC191.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204792; PubMed=9535817;
RA Tanaka N., Ohuchi N., Mukai Y., Ozaka Y., Ohtani Y., Tabuchi M.,
RA Bhyuyuan M., Fukui H., Harashima S., Takegawa K.;
RT "Isolation and characterization of an invertase and its repressor
RT genes from Schizosaccharomyces pombe.";
RL Biochem. Biophys. Res. Commun. 245:246-253(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,  
RA Welte J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritsch C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of *Schizosaccharomyces pombe*,"  
RN Nature 415:871-880(2002).  
RN [3]  
RP SEQUENCE OF 88-581 FROM N.A.  
RC STRAIN=PR745;  
RX MEDLINE=90253381; PubMed=2187435;  
RA Moreno S., Sanchez Y., Rodriguez L.;  
RT "Purification and characterization of the invertase from  
RT *Schizosaccharomyces pombe*,"  
RT Biochem. J. 267:697-702(1990).  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-  
CC fructofuranoside residues in beta-D-fructofuranosides.  
CC -!- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS 67%. THIS IS  
CC COMPOSED OF EQUIAMOLAR AMOUNTS OF MANNOSE AND GALACTOSE. THERE IS  
CC ALSO A SMALL AMOUNT OF GLUCOSAMINE PRESENT.  
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB011433; BAA25684.1;  
DR EMBL; AL049644; CAB41057.1;  
DR EMBL; DB9242; BAA13903.1;  
DR InterPro: IPR001362; GH\_32.  
DR Pfam: PF00251; Glyco\_hydro\_32; 1.  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Hydrolase; Glycosidase; Glycoprotein; signal.  
FT SIGNAL 1 22  
FT CHAIN 23 581  
FT ACT\_SITE 97 97  
FT BY\_SIMILARITY.  
FT CARBOHYD 37 37  
FT CARBOHYD 40 40  
FT CARBOHYD 46 46  
FT CARBOHYD 57 57  
FT CARBOHYD 62 62  
FT CARBOHYD 79 79  
FT CARBOHYD 168 168  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT N-LINKED (GLCNAC. . .) (PROBABLE).  
FT MISSING (IN REF. 3).  
FT H -> Y (IN REF. 3).  
FT A -> P (IN REF. 3).  
FT Q -> L (IN REF. 3).  
FT D -> N (IN REF. 3).  
FT I -> M (IN REF. 3).  
FT M -> L (IN REF. 3).  
FT S -> P (IN REF. 3).  
SQ SEQUENCE 581 AA; 64407 MW; 70206A6CD1F27EC4 CRC64;

Query Match 12.3%; Score 348; DB 1; Length 581;

Best Local Similarity 24.9%; Pred. No. 7.9e-18;

Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;

QY 28 PPAARASAPGSL-----RAVYHMTFPGSGWLCDPQRPVTHGAYQLYY 69  
DB 53 PPFVNTAPNGTCLNGYNEPLPSGYNATDRPKIHFTSPSGMNDPNGLVYTGGVYHMF 112  
QY 70 LHSQD--NNGPGGDHASTTGDVAF-----THHGTVMPLRDPDPVWMSGVAVGTA 117  
DB 113 QVSPKTLTAGVHWGHTVSKDLHWNYPYAIYDPDEHENGVLSL-----PFGSAVVDVH 167  
QY 118 NTAGFGAG-----AVVALATQPTDGVRYQYQYLYWSTDGGFTFTALPDVPIVNTDGRA 171  
DB 168 NSSGLFSNDTPEERIVLYTDHWTGVA--ERQAIYATTDGGYTFK-----KY 213  
QY 172 ATPAEIENAEWFRDPKIHWDTARGCWCVIGELRY--AAFTSPNLRDWTLRNFDYPN 229  
DB 214 SGNPVLIDINSLOFRDPKIVDFDANRWVMIVAMSONYQAFYSSYDLIHWLSTLSEVSTG 273  
QY 230 HALGGIECPDLFEITADDGTRH--WLAASMDAYGIGLPM---TYAYWTGTWDEQEQHADD 285  
DB 274 YLGLOVECPDMARVPVEGTDEYKWLFIINP---CAPLGGSVVQYFVGDNWGTNEVPDD 330  
QY 286 LPQWLDGWDYAAVTPSIDAPETKRIAIAMNNWKAARDVPTDASLTYNGQNSIVR 345  
DB 331 GQTRFVDLGDKEYASALYHSSA--NADVICGVGWSNMQY--TQAPQV---FRSANTVAR 385  
QY 346 ELRLARQPGGWYT---LLSTP--VAALTN--YVTATTLTPDKTVDSAVLPWNGRAYE- 396  
DB 386 KTLRDVDPONPMNTLSLOTPLNLSLLRDELFTAPVINSSSSLSGSPITLPSNTAFEF 445  
QY 397 -IETDIAMDTATNVGISVGR---SPDGTR---HTNIGKYGADLYVDYRGPDSLGY- 446  
DB 446 NVTLSTINYTEGCTGTCGLGRIIIDSDDPYLQGISVDVDFAASTLVINRAKAGWGFNSL 505  
QY 447 APYSRAAADPDGARSVHRLIILVQTSVEFVNAGHTVLSQQVHFAEGDT 496  
DB 506 FTFPSANDIYIYG--NVTLYGIVDNGLLLEYVNGEKTYYTNDFFFLQAGT 553

RESULT 10

ID RAFT\_ECOLI STANDARD; PRT; 476 AA.

AC P16553;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DE 01-MAR-1992 (Rel. 21, Last annotation update)

GN Raffinose invertase (EC 3.2.1.26) (Invertase).

OS RAFT.

OG Escherichia coli.

OC Plasmid PRSD2.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.



```
Db 173 -SSNQFDPKVFHFKSMOHCSE-----IARVKIQIF-GSANLKNWVLSNFSSGYG 226
QY 230 HALGGIECPDLFEITADDGTR-HWVLAASMDAYGIGLPM---TYAYWTGTWGDGEFHADD 285
Db 227 NOYQ---MSRLIEVPIENSDSKWVFLAIP---GSLPGSGINQYFVGDGFQFVPPD 280
QY 286 LTPQWLDGWDAVYAAVTPSIDAPEKRLAIAWMNWKYAAVDPTDASDYGNGQNSIVR 345
Db 281 SQRFVDIGKDFYAFQTFSEV---EHGVLGLAWASNQYADQ-VPTNP---WRSSTSLAR 333
QY 346 ELRLARQPGGWYLLSTPVA---LTYVTATTTLPDRTDGGSAVLP---WNGRAY 395
Db 334 NYTLR-----YVIOMLKLANTDKSVLPDSINVVDKLLKNVKNKPIKTNFKGSG 387
QY 396 EIELDIAMDATNWSVGR-----SPDGRHRTNIGKYGAD-----LYVDRG----- 437
Db 388 LFDENITF-KVLNLNVPKGTHTPDLINSOELNSSVDSIKIGFSSOSLFIIDRHIPNVE 446
QY 438 ----PSDLACYSLAPYSRAAPIDPGARSVHLRLVDYTSQVEFVNAGHTVLSQOVH 490
Db 447 FPKQFFTDKLAAY-LEPLD-----YDQDLRVFSLGYIVDKNIIELYFNDGTVMNTTF 500
QY 491 FAEQ-----DTGISLYTDGGFAHFTGIVVREIGQ 519
Db 501 MGEGKYPHDIQIYVDTPEELFELESVIRELNK 533

RESULT 12
INVA_MAIZE STANDARD; PRT; 590 AA.
AC P49174;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Beta-fructofuranosidase, cell wall isozyme precursor (EC 3.2.1.26)
DE (Sucrose-6-phosphate hydrolase) (Invertase).
OS Zea mays (Maize)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Black Mexican Sweet;
RA Shanker S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -! SUBCELLULAR LOCATION: Cell wall.
CC -! INDUCTION: By wounding and bacterial infection.
CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17695; AAA64487.1; -
CC MaizeDB; 113032; -
CC InterPro; IPR001362; GH_32.
CC Pfam; PF00251; Glyco_hydro_32; 1.
CC PROSITE; PS00609; GLYCOSYL_HYDROL_F32; FALSE NEG.
KW Hydrolase; Glycosidase; Glycoprotein; Cell wall; Signal.
FT SIGNAL 1 28 POTENTIAL..
FT CHAIN 29 590 BETA-FRUCTOFURANOSIDASE, CELL WALL
FT ACT_SITE 68 68 ISOZYME.
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 590 AA; 65198 MW; 29645561E99B892 CRC64;
```

```
Query Match 11.0%; Score 312; DB 1; Length 590;
Best Local Similarity 25.3%; Pred. No. 3.2e-15;
Matches 150; Conservative 81; Mismatches 211; Indels 150; Gaps 34;

QY 19 LALIFGGA-----VPPAARASAPGS-----LRVYVHMTTPPSGWLCDPQRPVTHGAYO 66
Db 21 LAURLAGASHVHRSLEAEAPSPASIVSPLLRTYGHFQPPMNWINDNAPLYYKGWTH 80
QY 67 LYLHSDQNNPGG--WDHASTTGDGVAFTH-HGTVMPLRPD--FPVWSGSAVVGTTANTAG 121
Db 81 LFQYYPKGAWGNIVWAHSVSRDLINWVALEPAIYPSIDKYGCMGSGSATI-----LE 135
QY 122 FGAGAVVALATQDGVRYQEOLYKWSDDGGFTTALP-----DPVIVNTD-----GRAA 172
Db 136 DGTPALYTGIDRAD--INVQOVL-----ALPKDASDPLLRWEKPEEYNPA 182
QY 173 TTPAETENAEWPRDKIHWDTARGENVCVIGRLR---YAAFTSNLNDWTLRRNFDYP 228
Db 183 TPAAGGINATQFHDPTTAMRHA-GHWRMLVGSVGRGARGMALVYRSRDFRKWTAK---HP 238
QY 229 NH--ALGGI-ECPLDFEI-----TADGTRHWVLAASMDAYGIGLPMTYAYWT-G 274
Db 239 LHSAAITGMWECDFPFVSPGLOAGLDTAPGCK-YVLKSSLDL-----TRYDIYTG 291
QY 275 TWDG---EQFHADDLTPQW-----LDWGMWYAAVTPSIDAPETKRLAIAMNNWYAAAR 327
Db 292 SYDGGKDRYPPDPAGDYHRRRYDYG-NYASKTF--YDPVERRRVLLGWANE-----SD 344
QY 328 DYPDTASDGYNGNSIVRELRLARQPGWYTLTSTPVAALTNYVTATTTLPDPTVDSAV 387
Db 345 SVPDDRAKAGWAGIHAIPRKIWL--DPTG-KQLLOWIHEVEKLRKAV-----SVDKLV 396
QY 388 LPWNG-----RAYETELDIAMDTATNVGISV-----GRSPDGTTRHTNIGYGDLYV 434
Db 397 KPGDHEFVTGIATYQADVEVSFELEAGTSLEKAEAFDPAVDDDAQKLCGVKGADARG 456
QY 435 DRGP-----SDLAGYSLAP--YSRAAA---P 455
Db 457 GVGPGFLWLASADLOERTAVFRVFRDGHGPKVKVLMCTDPTKSSLSPLDYKPTFAGFVD 516
QY 456 IDPGARSVHLRLVDYTSQVEFVNAGHTVLSQOVH--FAEG-DTGISLYTDG 504
Db 517 ADISSGKITLRSLIDRSVVSFSGAGGKTCILSRVPSIAGVDAHLVYFNNNG 568

RESULT 13
INVA_ZYMMO STANDARD; PRT; 512 AA.
AC P35636;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Sucrose-6-phosphate hydrolase E1 (EC 3.2.1.26) (Sucrase E1)
DE (Invertase E1).
GN INVA.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Z6C;
RX MEDLINE=91291345; PubMed=1368686;
RX Yanase H., Fukushi H., Ueda N., Maeda Y., Toyoda A., Tonomura K.;
RT "Cloning, sequencing, and characterization of the intracellular
RT invertase gene from Zymomonas mobilis.";
RL Agric. Biol. Chem. 55:1383-1390(1991).
CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
```



```

Db 65 SPAPSGVSGQSDKTRFDVAGASHVSYAWSNMLISWQRTAYHFQPKNMNDPGLY 124
QY 61 THGAYOLYLHSDQNN--GPGGDHASTTGVAFTHHGTVP--LRPD-----FPVWSGS 111
Db 125 HKGMYHLFYQYNPDSAINGNITWGHAVSKDLI---HWLYLPFAMVPDOWYDINGVWTGS 180
QY 112 AVVGTANTAGFAGAVVALATQPTDGVKRYQOYLYWSTDGGFTTALPDPIV---VNTD 168
Db 181 ATI-----LPDQIMLMYLTGTDY--VQVQNLAYPAN-----LSDPLLLDWMVKF 224
QY 169 GRAATPAEIAENAEWRDPKIH--DTARGEWCV-----IGRLRYAAFYTSNRLDWTLLR 223
Db 225 GNPVLVPPPGIGYKDFRPTTAWTGPQNGWLLTIGSKIGKTGVALVYETSNTSKL-- 282
QY 224 NFDPYPHALGGI---ECPDLFEITA-----DDGTRHWVLAASM-----DAYGIG 264
Db 283 -LDGVLHVPCTGMWECVDFEYPVSTKTNGLDTSYNGPGVKH-VLKASLDNNKQDIYAIG 340
QY 265 LPMYAYWTGTWDEQFHADDLPQWLDGWDWYAAVTPWPSIDAPETKRLAIAMNNWKY 324
Db 341 ---TYDLGNKMKTPDNPDELDCGIGLADYG-KYYASKTE--YDPKKERRVLCWIGETDS 394
QY 325 AARDVPTDASDGYNGONSIVRELRLARQCGWYTLTSTPVAALTNVYATITLTPDRTVD- 383
Db 395 ESADL-----QKQWASVQSIPRTVLYDKKTG--THLLOWPVEIESLRVGDPTV--KQVDL 446
QY 384 -GSAVLPWNGRAYEIELOIADMT-----ATNVGISVGRSPDGTGTRHTNIGRYG-- 429
Db 447 QPQSIELLRVDSAEILDIEASFVDKVALQGIIEADHVGFSCSTSGAASRGILGPGVI 506
QY 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAIPDPGARS 462
Db 507 VIADQTLSELTPYFYISKAGDGRAETHFCADQTRSEAPGVGKQVYGSVPVLDGEKHS 566
QY 463 VHLRIILVDTSQSVFVNAGHTVLSQOVH 490
Db 567 --MRLLVDHSIVESFAQGGRTVITSRIY 592

RESULT 15
INVL_CAPAN STANDARD; PRT; 640 AA.
AC P3761;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Acid beta-fructofuranosidase AIV-18 (EC 3.2.1.26) (Acid sucrose-6-
DE phosphate hydrolase) (Acid invertase).
OS Capsicum annuum (bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fruit;
RA Choi D., Lee K.-W., Kim S.;
RT "Isolation and characterization of acid invertase cDNA clone in Hot
RT pepper (Capsicum annuum L.) fruits.";
RL J. Plant Biol. 40:298-303(1997).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: U87849; AAB48484.1; --
DR InterPro: IPR001362; GL_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
DR PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Transmembrane; Glycoprotein.
FT TRANSMEM 34 54 POTENTIAL.
FT ACT_SITE 126 126 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 640 AA; 70620 MW; D3G628B7A7E6870B CRC64;

Query Match 9.2%; Score 260.5; DB 1; Length 640;
Best Local Similarity 23.8%; Pred. No. 1.e-11;
Matches 138; Conservative 77; Mismatches 221; Indels 143; Gaps 32;

QY 1 MPAISERRAVLQAGAGALALIFGGAVPPAARASAP-GSLRAYHYHTPPSGWLCDPQRPV 59
Db 72 LTPATSRGVSGVSEKTFKDVSGTSQVSYTWSNMLNMQRTAYHFQPKNMNDPGL 131
QY 60 TTHGAYOLYLHSDQNN--GPGGDHASTTGVAFTHHGTVP--LRPD-----FPVWSG 110
Db 132 YHKGWHLFYQYNPDSAINGNITWGHAVSTDLI---HWLYLPFAMVPDOWYDINGVWTG 187
QY 111 SAVVGTANTAGFAGAVVALATQPTDGVKRYQOYLYWSTDGGFTTALPDPIV---VNT 167
Db 188 SATI-----LPDGLIMLYLTGTDY--VQVQNLAYPAN-----LSDPLLLDWMVKY 231
QY 168 DGRAATPAEIAENAEWRDPKIH--DTARGEWCV-----IGRLRYAAFYTSN--LRDWT 220
Db 232 QGNPVLVPPPGIGYKDFRPTTAWTGPQNGWLLTIGSKIGKTGVALVYETSNTSKL 291
QY 221 LRRNFDPYPHALGGI---ECPDLFEITADD-----GTRHWVLAASMDAYGIGLP 266
Db 292 L-----HAPGTGMWECVDFEYPVSTLDANGLDTSYNGPGIKH-VLKASLD--N 337
QY 267 MTYAYWTGTD--GEQFHADDLPQW-----LQWGDWYAAVTPWPSIDAPETKRLAIA 317
Db 338 KODHYVIGTYDVPKNKFSPDN--PDLDCGIGLRLDYG-KYYASKTE--YDPKKORRVLMG 392
QY 318 WNNWKYAADVPTDASDGYNGONSIVRELRLARQCGWYTLTSTPVAALTNVYATITTL 377
Db 393 WIGETDSESADL-----QKQWASVQSIPRTVLYDKKTG--THLLOWPVEIESLRSDPKV 446
QY 378 PDR-----VDGSAVLPWNGRAYEIELOIADMTAT-----NVGISVGRSPDGT 420
Db 447 KEVNLOPGSIELLHVDCAA-----QFDIEASFVDRVTLGIIIEADVGYNCSTSGGAA 499
QY 421 RHTNIGKYG-----AD-----LYVDRG-----PSDLAGYSLAPYSR 451
Db 500 SRGILGPGCVVVIADQTLSELTPYFYISKAGDGRAETHFCADQTRSEAPGVAKQVYG- 558
QY 452 AAAPIDPGARSVHLRIILVDTSQSVFVNAGHTVLSQOVH 490
Db 559 SSVPLVDGEKH-RMRLVDHSIVESFAQGGRTVITSRIY 596

```

Search completed: May 6, 2003, 12:32:41  
Job time : 12.4798 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:29:32 ; Search time 56.8904 Seconds  
(without alignments)  
1879.728 Million cell updates/sec

Title: US-09-868-328B-4  
Perfect score: 2837  
Sequence: 1 MTPAISRRRAVLQAGAGALA.....LYTDGPAHFTGIVVREIGQ 519

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues ,

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_21.\*
- 1: sp-archaea.\*
  - 2: sp-bacteria.\*
  - 3: sp-fungi.\*
  - 4: sp-human.\*
  - 5: sp-invertebrate.\*
  - 6: sp-mammal.\*
  - 7: sp-mhc.\*
  - 8: sp-organelle.\*
  - 9: sp-phage.\*
  - 10: sp-plant.\*
  - 11: sp-rodent.\*
  - 12: sp-virus.\*
  - 13: sp-vertebrate.\*
  - 14: sp-unclassified.\*
  - 15: sp-rvirus.\*
  - 16: sp-bacteriap.\*
  - 17: sp-archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----|-------------|
| 1          | 2837   | 100.0         | 521    | 2  | Q9KJD0      |
| 2          | 2352   | 82.9          | 517    | 2  | O50585      |
| 3          | 2333   | 82.2          | 530    | 2  | Q9EV09      |
| 4          | 1047.5 | 36.9          | 620    | 2  | Q93R69      |
| 5          | 944    | 33.3          | 516    | 16 | O07003      |
| 6          | 880.5  | 31.0          | 514    | 16 | O07003      |
| 7          | 659.5  | 23.2          | 395    | 2  | P94469      |
| 8          | 562.5  | 19.8          | 512    | 2  | Q45372      |
| 9          | 527.5  | 18.6          | 501    | 2  | Q9LAL1      |
| 10         | 515    | 18.2          | 622    | 2  | Q45155      |
| 11         | 513    | 18.1          | 534    | 2  | Q9RBJ1      |
| 12         | 486.5  | 17.1          | 485    | 2  | Q9RLU2      |
| 13         | 469.5  | 16.5          | 493    | 2  | Q9APY5      |
| 14         | 468.5  | 16.5          | 493    | 2  | Q93T55      |
| 15         | 467.5  | 16.5          | 943    | 2  | Q44109      |
| 16         | 457    | 16.1          | 537    | 3  | O42801      |

|    |       |      |      |    |        |
|----|-------|------|------|----|--------|
| 17 | 448   | 15.8 | 537  | 3  | Q96TU3 |
| 18 | 433   | 15.3 | 516  | 3  | O74641 |
| 19 | 427   | 15.1 | 516  | 3  | O74642 |
| 20 | 427   | 15.1 | 516  | 3  | O94220 |
| 21 | 425   | 15.0 | 515  | 3  | O9HFA5 |
| 22 | 423   | 14.9 | 516  | 3  | Q96W28 |
| 23 | 422.5 | 14.9 | 550  | 5  | O02490 |
| 24 | 421   | 14.8 | 516  | 3  | Q8X217 |
| 25 | 416.5 | 14.7 | 533  | 3  | O94224 |
| 26 | 401.5 | 14.2 | 1142 | 16 | O97J24 |
| 27 | 389.5 | 13.7 | 515  | 3  | O00056 |
| 28 | 386   | 13.6 | 750  | 2  | O31411 |
| 29 | 366.5 | 12.9 | 555  | 3  | Q9UWF4 |
| 30 | 356   | 12.5 | 477  | 2  | O86076 |
| 31 | 355   | 12.5 | 477  | 16 | O8XC10 |
| 32 | 354.5 | 12.5 | 609  | 3  | Q9Y746 |
| 33 | 320.5 | 11.3 | 591  | 10 | Q9ZTL2 |
| 34 | 313.5 | 11.1 | 453  | 3  | O42878 |
| 35 | 304   | 10.7 | 513  | 5  | Q9U0Y9 |
| 36 | 304   | 10.7 | 670  | 10 | Q8RVK8 |
| 37 | 301   | 10.6 | 581  | 5  | Q9U0Y8 |
| 38 | 301   | 10.6 | 812  | 2  | Q9RBA5 |
| 39 | 297.5 | 10.5 | 581  | 10 | Q43856 |
| 40 | 291   | 10.3 | 695  | 3  | Q9C214 |
| 41 | 283.5 | 10.0 | 580  | 10 | Q43799 |
| 42 | 283   | 10.0 | 650  | 10 | O42722 |
| 43 | 283   | 10.0 | 661  | 10 | O94C05 |
| 44 | 279   | 9.8  | 636  | 10 | O94114 |
| 45 | 279   | 9.8  | 646  | 10 | Q8RVH4 |

ALIGNMENTS

RESULT 1

| ID | Q9KJD0   | PRELIMINARY: | PRT: | 521 AA. |
|----|--|--------------|------|---------|
| AC | Q9KJD0;  |              |      |         |
| DT | 01-OCT-2000 (Tremblrel. 15, Created)                             |              |      |         |
| DT | 01-OCT-2000 (Tremblrel. 15, Last sequence update)                |              |      |         |
| DT | 01-MAR-2002 (Tremblrel. 20, Last annotation update)              |              |      |         |
| DE | Levan fructotransferase.   |              |      |         |
| GN | LFTA.  |              |      |         |
| OS | Arthrobacter ureafaciens.  |              |      |         |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;          |              |      |         |
| OC | Actinomycetales; Micrococccineae; Arthrobacter.                  |              |      |         |
| OX | NCBI_TaxID=37931;  |              |      |         |
| RN | [1]  |              |      |         |
| RP | SEQUENCE FROM N.A.   |              |      |         |
| RC | STRAIN=K-2032;   |              |      |         |
| RA | Song K.B., Rhee S.K., Yoo E.J.;                                  |              |      |         |
| RT | "Nucleotide sequence of levan fructotransferase gene (lfta) from |              |      |         |
| RT | Arthrobacter ureafaciens K-2032."                                |              |      |         |
| RL | Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.          |              |      |         |
| CC | -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.     |              |      |         |
| DR | EMBL; AF181254; AAF3829.1;                                       |              |      |         |
| DR | InterPro: IPR001362; GH_32.                                      |              |      |         |
| DR | Pfam: PF00251; Glyco_hydro_32; 1.                                |              |      |         |
| KW | Glycosidase; Hydrolase; Transferase.                             |              |      |         |
| SQ | SEQUENCE 521 AA; 56590 MW; 863F3FFC4622D2B3 CRC64;               |              |      |         |

Query Match 100.0%; Score 2837; DB 2; Length 521;  
Best Local Similarity 100.0%; Pred. No. 4.9e-179;  
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |    |   |     |
|----|----|---|-----|
| QY | 1  | MTPAISRRRAVLQAGAGALALIFGGAVPPAARASAPGSLRAVYHMTPTPSGWLCDPQRPVT | 60  |
| Db | 1  | MTPAISRRRAVLQAGAGALALIFGGAVPPAARASAPGSLRAVYHMTPTPSGWLCDPQRPVT | 60  |
| QY | 61 | THGAYQLYLLHSDQNGPGWDHASTTDCGVAFTTHGTVMLPRPDPFPVWSGSAVVGTA     | 120 |
| Db | 61 | THGAYQLYLLHSDQNGPGWDHASTTDCGVAFTTHGTVMLPRPDPFPVWSGSAVVGTA     | 120 |

```
QY 121 GFAGAVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTPAEIN 180
Db 121 GFAGAVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTPAEIN 180
QY 181 AEFRRDPKIHWDATARGWCVCIGRLRYAAYTSPNLRDWTLRNFDPNHALGGIECPDL 240
Db 181 AEFRRDPKIHWDATARGWCVCIGRLRYAAYTSPNLRDWTLRNFDPNHALGGIECPDL 240
QY 241 FEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDGWDWYAA 300
Db 241 FEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDGWDWYAA 300
QY 301 VTWPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360
Db 301 VTWPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360
QY 361 STPVAALTNVYATTTLPDRTPDGSVAVLPWNGRAYEIELDIANDTATNVCISVGRSPDGT 420
Db 361 STPVAALTNVYATTTLPDRTPDGSVAVLPWNGRAYEIELDIANDTATNVCISVGRSPDGT 420
QY 421 RHTNICKYGADLYVDRGPSLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVVEFVNA 480
Db 421 RHTNICKYGADLYVDRGPSLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVVEFVNA 480
QY 481 GHTVLSQVHFAGSDGTGISLYTDGGAHFTGIVVREIGQ 519
Db 481 GHTVLSQVHFAGSDGTGISLYTDGGAHFTGIVVREIGQ 519

RESULT 2
ID O50585 PRELIMINARY; PRT; 517 AA.
AC O50585;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Levan fructotransferase.
OS Arthrobacter nicotinovorans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Micrococccaceae; Arthrobacter.
OX NCBI_TaxID=293320;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98101974; PubMed=9438987;
RA Saito K., Yokota A., Tomita F.;
RT "Molecular cloning of levan fructotransferase gene from Arthrobacter
RL nicotinovorans GS-9 and its expression in Escherichia coli.";
RL Biosci. Biotechnol. Biochem. 61:2076-2079(1997).
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AB001984; BAA24915.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PR00251; Glyco_hydro.32; 1.
KW Glycosidase; Hydrolase; Transferase.
SQ SEQUENCE 517 AA; 56509 MW; D3563EF9E52A1FE3 CRC64;

Query Match 82.9%; Score 2352; DB 2; Length 517;
Best Local Similarity 81.3%; Pred. No. 4.5e-147;
Matches 421; Conservative 41; Mismatches 52; Indels 4; Gaps 2;

QY 1 MTPAISRRVLOGAGAGALALIFGGVPPAARASAPGSLRAYVHMTPPSGWLCDDPQPVY 60
Db 1 MTYDISRRVLOGAGAGALALFMSNIPVAAHAQA--SURIYHMTTPPSGWLCDPQPVH 58
QY 61 THGAYQLYLHSDQNGPGGWDHASTTGDVAFTHHGTVMPLRPDPFVWSSAVVGQTANTA 120
Db 59 TNGAYQLYLHSCQNGPGGWDHATTGDCGVSYTHGHVMPMQDPFVWSSAVVDVANTA 118
QY 121 GFAGAVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAATTPAEIN 180
Db 119 GFAGAVVALATOPTDGV--KFGQEQYLWSTDDGYSFTALPDVPVINTDGRATTAPAEV 176
QY 181 AEFRRDPKIHWDATARGWCVCIGRLRYAAYTSPNLRDWTLRNFDPNHALGGIECPDL 240
Db 181 AEFRRDPKIHWDATARGWCVCIGRLRYAAYTSPNLRDWTLRNFDPNHALGGIECPDL 240
```

```
Db 177 AEFRRDPKIHWDATARGWCVCIGRLRYAAYTSPNLRDWMOWKSNFDYPNHALGGIECPDL 236
QY 241 FEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLDGWDWYAA 300
Db 237 FEMTAGDTRHWWVFGASMDAYSIGLPMYAYWTGSWNGTAFIADNLTPQWLDGWDWYAA 296
QY 301 VTWPSIDAPETKRLAIAMNNWKYAAARDVPTDASDGYNGQNSIVRELRLARQPGWYTL 360
Db 297 VTWPAVEAPETKRLAIAMNNWKYAAARNVPTDASDGYNGQNSITRELRLERQSGWYTL 356
QY 361 STPVAALTNVYATTTLPDRTPDGSVAVLPWNGRAYEIELDIANDTATNVCISVGRSPDGT 420
Db 357 STPVPALSNYATSTTLPDRTPVNGSFVLPWNGRAYEIELDISWDTAANVGVSGRSDGS 416
QY 421 RHTNICKYGADLYVDRGPSLAGYSLAPYSRAAAPIDPGARSVHLRLILVDTQSVVEFVNA 480
Db 417 RHTNICKYGDELIVDRASSEQSYALAPYTRAAPIDANARSVHLRLIFVDTQSVVEFVNS 476
QY 481 GHTVLSQVHFAGSDGTGISLYTDGGAHFTGIVVREIG 518
Db 477 GHTVWSQVHFAGSDGTGISLYADGGPANFTGITIREFG 514

RESULT 3
ID Q9EVQ9 PRELIMINARY; PRT; 530 AA.
AC Q9EVQ9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Levan fructotransferase.
GN LFTM.
OS Microbacterium sp. AL-210.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=135249;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL-210;
RA Yang S.J., Cha J.;
RT "Levan fructotransferase gene (lftm) from Microbacterium sp. AL-210.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AF294730; AAG01554.1; -
DR InterPro; IPR001362; GH_32.
DR Pfam; PR00251; Glyco_hydro.32; 1.
KW Glycosidase; Hydrolase; Transferase.
SQ SEQUENCE 530 AA; 57458 MW; 1A7E78D19EID0596 CRC64;

Query Match 82.2%; Score 2333; DB 2; Length 530;
Best Local Similarity 80.2%; Pred. No. 8.2e-146;
Matches 421; Conservative 36; Mismatches 56; Indels 12; Gaps 2;

QY 4 AISRRAVLOGAGAGALALIFG-----GAVPAARASAPGSLRAYVHMTPPSGWLC 53
Db 5 SFSRMTLKGTSVGAALLIGSSPTLAQSSAGLRAAAAAA--GSRAYVHMTTPPSGWL 62
QY 54 DPORPVTHGAYQLYLHSDQNGPGGWDHASTTGDVAFTHHGTVMPLRPDPFVWSSAV 113
Db 63 DPORPVTHGAYQLYLHSCQNGPGGWDHATTGDCVAFTHHGTVMPLRPDPFVWSSAV 122
QY 114 VGTANTAGGAVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAAT 173
Db 123 VGTANTAGGAVVALATOPTDGVRRYQEQYLWSTDDGGFTTALPDVPVINTDGRAT 182
QY 174 TPAEINAEWFRDPKIHWDATARGWCVCIGRLRYAAYTSPNLRDWTLRNFDPNHALG 233
Db 183 TPAEINAEWFRDPKIHWDALRGWVCVIGRARIYAFYTSNLRDQWTSNFDYPNHALG 242
QY 234 GIECPDLFEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWDEGFHADDLTPOWLD 293
Db 243 GIECPDLFEITADGTRHWWLAASMDAYGIGLPMYAYWTGTWNGTHFIAGNLVPOWLD 302
```

```
QY 294 GNDWTAAYTWPISIDAPETKRLAIAMNNWKYAARDVPTDASDGYNGONSIVRELRARQP 353
Db 303 GNDWTAAYTWPRAEAPETKRLAIAMNNWKYAARDVPTDASDGYNGONSITRELRARQE 362
QY 354 GGYWTLTSPVAALTNVYATTTLPDRTYDGSNVLFPWNGRAYEIELDIANDTATNVGISV 413
Db 363 GGYWTLTSPVAALTNVYATTTLPDRTYDGSNVLFPWNGRAYEIELDIANDTATNVGISV 422
QY 414 GRSPDGTHTNIGKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPCARSVHLRLVDTOS 473
Db 423 GRSADGARHTNIGKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPCARSVHLRLVDTOS 482
QY 474 VEVFNAGHTVLSQOVHFAEGDTGSLYTDGGPAHFTGIVVREIG 518
Db 483 VEVFNAGHTVLSQOVHFAEGDTGSLYTDGGPAHFTGIVVREIG 527

RESULT 4
QY3R69
ID Q93R69 PRELIMINARY; PRT; 620 AA.
AC Q93R69;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Levanase.
OS Microbacterium laevaniformans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Micrococcales; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=36807;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=ATCC 15953;
RX Song E., Cha J.;
RA "Microbacterium laevaniformans levanase-complete cds.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AB062062; BAB59060.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 620 AA; 67667 MW; 82B4909328CBAC2 CRC64;

Query Match 36.9%; Score 1047.5; DB 2; Length 620;
Best Local Similarity 45.3%; Pred. No. 5.2e-61;
Matches 243; Conservative 55; Mismatches 190; Indels 49; Gaps 16;

QY 3 PAI--SRRVLOG---AGAGALALIFGGVAPPAARASAPGSLRAYVHTMPTPSGWLCDPQR 57
Db 34 PALPARRALAGLALAGSAAVLAVG---PAAAEEDAP-SLRARYHFTVPDHWMDPQR 89
QY 58 PVTTHGAYQLYLH-SDQNGPGG-----WDHASTTDGVAFTHHGTVMF--LRPDPFVWS 109
Db 90 PVVIGDLVHLYLYNADYDANPRANFTGTEWRLATSADGVAFADQGVAAKPTNANYDLWS 149
QY 110 GSAAVGTANTAGGAGVALATQ---PTDGVV-----KYQGYLXWSTDGDTFTALPDP 162
Db 150 GSAAVVDHAGTAGGCAVVMVLTMDHPHTAAQKLDASGQAQFLWYSDVGGRTFRPDGDE 209
QY 163 VIVNTDGRAATTPAEIENAEWFRDKIHWDTARGEWCVIGRLRYAAFTSPNLRDWTLR 222
Db 210 PVIPDGRD-----FRDKPVVMDDERQWRVALIAERDRVSFTYSPDLHRWT-- 256
QY 223 RNFDYPNALGIECPDLPFEITADGTRHVLAAASMDAYGIGLPMYAYWTGTWDEQFH 282
Db 257 RGEVYVNGAGITGIECPDLFLRADGGTTHWYMGVANGATNEPATYAYWTGSPDGFSTFV 316
QY 283 ADDLPQMLDGDWNYAAVTPSIDAPETKRLAIAMNNWKYAARDVPTDASDGYNGONS 342
Db 317 PDVAAPQWLDHGFWDYAVTWEDPAAPLERLAIAMNNWDY-AHGAPWPRDGTGTDGTS 375
QY 343 IVRELRARQGGWTTLLSTPVAALTNVYATTTLPDRTYDGSNVLFPWNGRAYEIELDIA 402
Db 376 ITREIRLAR-AGAGYSLSPVADALQDHATHHTALGDVRVDGFTGLAYRGAVEITTRIS 434
```

```
QY 403 WDTATNVGISVGRSPDGTHTNIGKYGADLYVDRGSDLAGYSLAPYS---RAAAPIDPG 459
Db 435 RQADNVGQLRRSADSGSRHADAGLTRDAYLNRATG-----RPDSKVESRPLD-D 487
QY 460 ARSVHLRLVDTOSVEVFNAGHTVLSQOVHFAEGDTGSLYTDGGPAHFTGIVVRE 516
Db 488 ADTVELRLVDTTIEVFEVGDGRYTHSSQVFAFSGDGLALYTSGGPAVFRDLRITE 544

RESULT 5
QY07003
ID Q07003 PRELIMINARY; PRT; 516 AA.
AC Q07003;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein yveb.
GN YVEB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales.
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RP Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=168;
RX MEDLINE=98040433; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusier C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Katamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Taseuchi M., Tanakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Totsu V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler H., Wedler H., Weitzenegeer T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RC SEQUENCE FROM N.A.
RP STRAIN=168;
RA Kunst P., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1;
DR EMBL; Z99121; CAB15451.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase; Hypothetical protein; Complete proteome.
```

```
SQ SEQUENCE 516 AA; 58978 MW; EAEC526706C40762 CRC64;
Query Match 33.3%; Score 944; DB 16; Length 516;
Best Local Similarity 41.8%; Pred. No. 2.7e-54;
Matches 205; Conservative 71; Mismatches 176; Indels 38; Gaps 15;

QY 40 RAYVHMTTPSGWCLDQRPVTHGAYOLYLHS--DQNGPG--GWDHASTTGDGVAFTHHGT 97
DB 44 RAYVHMTTPSGWCLDQRPVTHGAYOLYLHS--DQNGPG--GWDHASTTGDGVAFTHHGT 103
QY 98 VMP--LRPDPFVMSGAVVGTANTAGFGAGVVALATQPTDGVKRYQEOYLWSTDGGFT 155
DB 104 AIPKYNPDGDIWGSVVDKENTAGFGKALVAIVTQPSAKDKK--QEOYLWSTDKGKS 162
QY 156 F-----TALPDPIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGEWVCVIGRLRYAA 209
DB 163 FKFSYGNVMPNP---GTDD-----FRDPKIVNDQDNKWMVMAEGSKIG 205
QY 210 FYTSPNLRDWTLRNFDYPNHALGGIECPDLFEITADDGTRHVVLAASMDAYGIGLPMYAYWVG 269
DB 206 FYSDNLKQHWHTSGF--FPEQA--GMVECPDLYMNRASDGTNKNWLGASANGKPGWKPNY 263
QY 270 AYVGTWGDGEPHADLLTPOWLDGWDWYAAVTPSIDA--PETKRLATAMNNKRYAAR 327
DB 264 AYVGTWGDGEPHADLLTPOWLDGWDWYAAVTPSIDA--PETKRLATAMNNKRYAAR 327
QY 328 DVPTDASGYNGQNSIVRELRLARQPGWYLLSTPVAALTNVYVATTTLPDRTVDGSAV 387
DB 323 NTPT--MKNFGNGTDSVIRELRLKEQ--DGYSLVSQPIEALEQLTVSTDEIQDVNGSKT 380
QY 388 LPNNGRAYEIELDIANDTATNVGISVGRSPDGRTHNIGKY--GADLYVDGRPSDLAGYS 445
DB 381 LSTGTGYOLDTDLSELKNAAGVRLRESEDDQKRHDVIGFAGGGYAYYNNRAATNPQDKS 440
QY 446 LAPYSRAAAPIDPGARSVHLRLILVDTQSVFVFNAGHTVLSQQVHPFAEGDTGISLYTDGG 505
DB 441 -NTYVESKAPYDNKRVKHLKILVDKTTIEVFVGDGKTVFSNEVFPKPEDKGITLYSDG 499
QY 506 PAHFTGIVVR 515
DB 500 TASFKNITVK 509

RESULT 6
Q97180 PRELIMINARY; PRT: 514 AA.
AC Q97180;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Levanase.
GN CAC1773.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79738.1;
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
KW Complete proteome.
SQ SEQUENCE 514 AA; 59636 MW; 50366383B2D97004 CRC64;
```

```
Query Match 31.0%; Score 880.5; DB 16; Length 514;
Best Local Similarity 40.7%; Pred. No. 4e-50;
Matches 200; Conservative 76; Mismatches 179; Indels 37; Gaps 14;

QY 40 RAYVHMTTPSGWCLDQRPVTHGAYOLYLHS--DQNGPG--GWDHASTTGDGVAFTHHGT 97
DB 45 REVVHTVTKKNDKQRPVFFDGEYHYLYLNKDPNGNGTEWROATSKDLTWKDEGV 104
QY 98 VMP--LRPDPFVMSGAVVGTANTAGFGAGVVALATQPTDGVKRYQEOYLWSTDGGFT 155
DB 105 CIPKYTNENGDIWGSFVVDQAQNTAGFGKALVAIVTQPSASMDK--QEOFLWYSTDRGT 163
QY 156 FTALPD--PVIWTDGRAATTPAEIENAEWFRDPKIHWDTARGEWVCVIGRLRYAAYTSP 214
DB 164 FKPSYQNPINP-----PCTKD-----FRDPKIIWDFKNNKVMVLAEGTKIGFESY 211
QY 215 NLRDWTLRNFDYPNHALGGIECPDLFEITADDGTRHVVLAASMDAYGIGLPMYAYWVG 274
DB 212 NLKNW--QITGDFFTNIGIVECPDIFMQSDNGNTYKWLGTSGANKVSGEPNTYAYWVG 269
QY 275 TWDGEOFHADDTLPQWLDGWDWYAAVTPSIDAPET--KRLAIAMNNKRYAARDVPTD 332
DB 270 NYDGKKFIADISTPKWLDYGDWYAAVTFESENOSKNLKKRYALAMNNWDY--ANNTPT- 327
QY 333 ASDGYNGQNSIVRELRLARQPGWYLLSTPVAALTNVYVATTTLPDRTVDGSAVLPWNG 392
DB 328 IQNFRGMDISVIREITLSKQKDNITSLSKPKRIENITTSIDQFKQISVKGIKHLKVOG 387
QY 393 RAVEIELDIANDTATNVGISVGRSPDGRTHNIGKY--GADLYVDGRPSDLAGYS LAP-- 448
DB 388 KAYQDSTDINWDKKNVGRRLRESLDKRRHDVIGFTEGKYSVNR-----AYTGNPK 441
QY 449 ---YSRAAAPIDPGARSVHLRLILVDTQSVFVFNAGHTVLSQQVHPFAEGDTGISLYTDGG 505
DB 442 SKKYVESRAPFDINNKVHLRFVDKVSVEFIDDKGITYSNEVFPRPEDKGITLFSIKG 501
QY 506 PAHFTGIVVREI 517
DB 502 KAVFKNITIKHI 513

RESULT 7
P94469 PRELIMINARY; PRT: 395 AA.
AC P94469;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan
DE fructanohydrolase).
GN SURC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12980;
RX MEDLINE=98007870; PubMed=9349714;
RA Li Y., Triccas J.A., Ferenci T.;
RT "A novel levansucrase-levanase gene cluster in Bacillus
RT stearothermophilus ATCC12980."
RL Biochim. Biophys. Acta 1353:203-208(1997).
CC -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 2,6-BETA-D-FRUCTO
CC FURANOSIDIC LINKAGES IN 2,6-BETA-D-FRUCTANS (LEVANS) CONTAINING
CC MORE THEN 3 FRUCTOSE UNITS.
CC -!- INDUCTION: IN CONTRAST WITH LEVANSUCRASE AND SUCRASE, LEVANSASE
CC SYNTHESIS IS NOT INDUCIBLE BY SUCROSE; NO INDUCER IS KNOWN FOR
CC LEVANSASE.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL: U34875; AAB97112.1;
DR InterPro: IPR001362; GH_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
```

KW Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 395 LEVANASE.  
FT ACT\_SITE 175 175 BY SIMILARITY.  
SQ SEQUENCE 395 AA; 44431 MW; 1AD4818A29F00EB CRC64;

Query Match 23.2%; Score 659.5; DB 2; Length 395;  
Best Local Similarity 41.2%; Pred. No. 1e-35;  
Matches 139; Conservative 51; Mismatches 114; Indels 33; Gaps 11;

Qy 85 STTDGVAFTHTGYNP--LRPDFVWGSAGVGTANTAGFAGAGAVVALATPTDGVKRYQ 142  
Db 35 SQTICWTWDGVAIPKYNPDGDIWTGVSVDKENTAGFGKNALVAITOPSAKDK-Q 93  
Qy 143 EOYLWSTDDGGFTF-----TALDDPVIVNTDGRAATPAEIEAENAEWRDPKIHWDTARG 196  
Db 94 EOYLWSTDDKCKSKFYSGNPMNP-----GTDD-----FRDPKVIWDDQDN 136  
Qy 197 EWVCVIGRLRYAAFTSPNLKRDWTLRRNFDYPNIALGIECPDLFEITADDGTRHWLAA 256  
Db 137 KWVMYMAEGSSIGFYESNLKDHWTSGF-EPEQT-GMVECPDLYMMRASDGTNKNWILGA 194  
Qy 257 SMDAYGIGLPMTYAYWTGTWDEGFHADDLTPOWLDDGWDKYAAVTPSIDA--PETKRL 314  
Db 195 SANGKPKGKPTIAYWTGSGFGKEFKADQTEAQLMDYGFWDYGGVTFEDSKSTDPLEKRY 254  
Qy 315 AIAMNNMKYAAARDVPTDASDYGNGQNSIVRELRLAROGGWYTLTTPVAALTNVYVAT 374  
Db 255 ALAMNNWDY-ANNTPT-MKNGFNGTDSVIREIRLKEQ-DCYISVSPQIEALEOLTYST 311  
Qy 375 TTLDPRTVDGSAVLPMWNGRAYEIELDIAMDPTATNVGI 411  
Db 312 DEIEDQDVNGSKTISITGDTYQLDLDLSWSSELKNAGV 348

RESULT 8  
Q45372 PRELIMINARY; PRT; 512 AA.  
ID Q45372  
AC Q45372  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Fructosyltransferase.  
GN LELA.  
OS Paenibacillus polymyxa (Bacillus polymyxa).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1406;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CF43;  
RX MEDLINE=94209215; PubMed=8157587;  
RA Bezzate S., Steinmetz M., Aymerich S.;  
FT "Cloning, sequencing, and disruption of a levanase gene of Bacillus polymyxa CF43.";  
RL J. Bacteriol. 176:2177-2183(1994).  
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
DR EMBL; Z26651; CA81392.1; -;  
DR InterPro; IPR001362; GH32.  
DR Pfam; PF00251; Glyco\_hydro\_32.1;  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Glycosidase; Hydrolase; Transferase.  
SQ SEQUENCE 512 AA; 58129 MW; A85C888ED80F8EF2 CRC64;

Query Match 19.8%; Score 562.5; DB 2; Length 512;  
Best Local Similarity 31.2%; Pred. No. 3.6e-29;  
Matches 162; Conservative 75; Mismatches 213; Indels 69; Gaps 18;

Qy 40 RAVVHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNNPG-----GWDHASTDGVAFTHH 95  
Db 13 RPQHYSPKEMNDPNGLVYFEGEYHLFYQHTPDTPDGFNMHWGHAVSKDLVHWTEL 72  
Qy 96 GTVMPLRPDPFVWGSAGVGTANTAGF-----GAGAVVALAT-----QPTDGVKRYQEQY 145

Db 73 PPAIPPGEDGAIFSGSAVVVDKNNNTSGFFDEEGSG-LVAITYTNEGKNSQP-----GKPOVQS 127  
Qy 146 LYWSTDDGGFTTALP-DPVIYNTDGRAATTPAEIEAENAEWRDPKTHWDTANGEWCVLGR 204  
Db 128 IAYSKDKGRTWKYEGNPVLPFTDLD-----FRDPKVIWDESSMTWMLAV 175  
Qy 205 LRYAAFTSPNLKRWTLRRNF--DYPNHALGGIECPDLFEITADD--GTRHWLAAAS-- 258  
Db 176 RDRVEFTSPNLKWSFASFGSDIPHIHRGIFECDFIRIOVDEDLNATTKWILMSVGD 235  
Qy 259 -----DAYGIGLPMTYAYWTGTWDEGEQFIADDL-----TPQWLDDGWDKYAAVT 302  
Db 236 RGVNPPDDPEPPAGSGM---MYFVGSDGKSETPDEALESIDTIKWVDYGSDFYAAVS 291  
Qy 303 WPSIDAPETKRLATAMNNMKYAAARDVPTDASDYGNGQNSIVRELRLAROGGWYTLTST 362  
Db 292 WNGISNEDGRKIWJGMMNNWRYAT---TLFSKEWKGRTSIPRELQRLTYPEG-LRLQT 346  
Qy 363 PVAALTNVYVATTTLPDRTV-DGSAVL-PWNGRAYEIELDIAMDPTATNVGISVGRSPDGT 420  
Db 347 PINELSQLRKPKTILSLQDLTIKPGMNVLSDISAAKAEIIAEFEIGTAVFEGFKVRKS--AN 404  
Qy 421 RHTNIGK--YGADLYVDRGPSDLAGYSLAPYSRAAADPGARSVHLRLVDTQSVFV 478  
Db 405 QETIIGYNISNEELFVDRTKSSATDFHSDFTAIHKATMKPHERIQLSIYLDWSSVFEV 464  
Qy 479 NAGHTVLSSQVHFABGDTGISLYTDGGAHFTGIVVREI 517  
Db 465 NHGKAILSDMIFPDPEESKGLLEYALGGELRVVSLQINDL 503

RESULT 9  
Q9LAL1 PRELIMINARY; PRT; 501 AA.  
ID Q9LAL1  
AC Q9LAL1  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Exo-inulinase.  
GN INU2.  
OS Pseudomonas mucidolens.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=46679;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kwon Y.-M., Choi Y.-J.;  
FT "DNA sequences and expression in Escherichia coli of an exo-inulinase gene (inu2) from Pseudomonas mucidolens.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
DR EMBL; AF129819; AAF44125.1; -;  
DR InterPro; IPR001362; GH32.  
DR Pfam; PF00251; Glyco\_hydro\_32.1;  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 501 AA; 57797 MW; 4043B86F20B4A4E7 CRC64;

Query Match 18.6%; Score 527.5; DB 2; Length 501;  
Best Local Similarity 31.3%; Pred. No. 7.1e-27;  
Matches 156; Conservative 72; Mismatches 223; Indels 47; Gaps 20;

Qy 40 RAVVHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNN--GPGWHDASTDGVAFTHHGT 97  
Db 11 RPQHFHTETWMDPNGLVYFEGEYHLFYQHPFGNTWGPWHCHAVSKDLIHWHD--L 68  
Qy 98 VMPLRPDF--PVGSGSAGVGTANTAGF--GAGAVVALATQP---TDGVRKYQEOYLWYST 150  
Db 69 PIALYPDHIGIFSGSAVIDNQDTSQFFNGGAGMVALTFHADKPPDPSGRPRQRLAYST 128  
Qy 151 DGGFTTALPDPIVNTDGRAATTPAEIEAENAEWRDPKIHWDTANGEWCVIGRLRYAA 210

```
Db 129 DRGRTH-----IKYEGNPFADAMTD---FRDPKVFVHKASNKWTMILIAAGNCVRI 177
QY 211 YTSPLNRDWTLRNFDPYPHALGGI-ECPLDFEITADDGTRH---WVLAASM---DAYGI 263
Db 178 YHSSNLKQKFASEFCAGNEGSHGTGVWECPLDFELPV-DGDRNRKQKVMVVSIGNSDEYLE 236
QY 264 GLPMTYAYTGTWGDGQF--HADDLTPQMLDWGWDWYAAVTPSIDAPETKRLAIAMNN 321
Db 237 G--SRTOYFIFGDFGIRFTNENDPETVLAWDHGRDNYAGVTWSIDPEEDGRRLFIAWMSN 294
QY 322 WKYAARDVPTDASDGYNGQNSIVRELRLARQPGWYLLSTPVAALTNVYVATTTLPDR- 380
Db 295 WKT-ANHTTPEI---WRSAMTIPRSLSRSTPEG-TRLFOTPVTELESURLEKLEWKDLE 349
QY 381 -TVDSAVLPWNGRAYEIELDIAMDATNVGISVGRSPDGRHTNIGKYGAD---LYVDR 436
Db 350 VTTNGVNLNISGDIIEIIAEFELNLALEFGFKVRSAD--QETIVG-YDAEQOQLFIDR 406
QY 437 GPSDLACYSLAPYSRAAAPIDPGARSVHLRLIVDTOSVEFVFNAGHTVLSQQVHFAEGDT 496
Db 407 TOSGVDFCEHFFCKRGAVMIPNQNIQHIFVDRSSVEFVGNNGELTMTDLIFPDSSST 466
QY 497 GISLYTDGGPAHFTGIWV 514
Db 467 GIEVYAKEGNVKLVSUML 484

RESULT 10
Q45155 PRELIMINARY; PRT; 622 AA.
AC Q45155;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Levanase.
GN SCRL.
OS Bacteroides fragilis.
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BF1;
RX MEDLINE=93259952; PubMed=8491724;
RA Blatch G.L., Woods D.R.;
RT "Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1."
RL J. Bacteriol. 175:3058-3066(1993).
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; M83774; AAA22924.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 622 AA; 70287 MW; 63424EB274CEB134 CRC64;

Query Match 18.2%; Score 515; DB 2; Length 622;
Best Local Similarity 30.2%; Pred. No. 6.4e-26;
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

QY 40 RAVYHMTFSGWCLDQRPVTHGAYQLYLLHSDQNN--GPGGWDHASTTDCGVAETHHCT 97
Db 133 RPLVHHTPLVGMNDANGLVYKDGETHLFYQYNPYGSMGNMHWGHSVSKDLVHWEH--- 189
QY 98 VMPLRDPDF-----VWGSAAVVGVTANTAGFGAGAVVALATOPTDGVKRYQQLYWSFD 151
Db 190 ---LEPALARDTLGHIFSGSVVDANTAGYAGAIYAFYTSASD--KNGQLOCHAYSTD 244
QY 152 GGTFTAL-PDPVYNTDGRAATPAEIAENAEWFRDPKTHWDTARGEWCVCVIGRLRYAAF 210
Db 245 NGRTETKYEKNPLVTPFDG-----LKDFRDPKVFVYAPDQKWMVVSADKEMRF 293
QY 211 YTSPLNRDWTLRNED-----YPNHALGGIECPDLFEITADDGTRH--WVLAASDAYGI 263
```

```
Db 294 YSENLEKWTYMSGWGEYGVQPSQ-----FECPDWVELPVDGNPDHKKWALLIVNVN---- 345
QY 264 GLPMTY-----AYTGTWGDGQFHADDL--TPQWLDWGDWYAAVTPSIDAPETKRLA 315
Db 346 --PGCYFGSATSQYFIFGDFGEKFCVCDNKPEITVKWLDWGDHYATVCFSTNG--DRTIA 400
QY 316 IAWNNWYKAAARDVPTDASDGYNGQNSIVRELRLARQPGWYLLSTPVAALTNVYVATTT 375
Db 401 VPMWSNQY-ANIVPTPRQ---FRSANALPRELSLYTQDGDY-MAAAPVEETKSLRKESR 455
QY 376 TLPDRIVDGS-----AVLPWNGRAYEIELDIAMDATNVGISVGRSPDGRHTNIGKYGAD 431
Db 456 EIPAFEGVAYHVDSLSDNKGAYEIELELAAGSABIMGLKLFNKGENVDIYISLPEKK 515
QY 432 LYVDRGPSDLA--GYSLAPYSRAA-----APIDPCARSVHLRLIV 469
Db 516 LYMDRTKSGIVDFGRDSAPHATEAHDRRKONSNIVYDDFALGTWAPVQR-AGNYKLDIFV 574
QY 470 DTQSEVFEVFNAGHTVLSQQVHFAEGDTGISLYTDGG 505
Db 575 DKCSVEIFLNGKIAMTNLIFFTPPYNQMSFYSRGG 610

RESULT 11
Q9RBJ1 PRELIMINARY; PRT; 534 AA.
AC Q9RBJ1;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Levanase precursor.
GN LSDB.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SRT4;
RX MEDLINE=96253999; PubMed=8704949;
RA Arrieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
RA Menendez C., Petit-Glatron M.F., Chambert R., Selman-Housein G.;
RT "Molecular characterization of the levansucrase gene from the
RT endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4."
RL Microbiology 142:1077-1085(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SRT4;
RX MEDLINE=99373266; PubMed=10441728;
RA Hernandez L., Arrieta J., Betancourt L., Falcon V., Madrazo J.,
RA Coego A., Menendez C.;
RT "Levansucrase from Acetobacter diazotrophicus SRT4 is secreted via
RT periplasm by a signal-peptide-dependent pathway."
RL Curr. Microbiol. 39:146-152(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SRT4;
RA Menendez C., Hernandez L., Mendoza M.F., Hevia P., Selman G.,
RA Arrieta J.;
RT "Molecular cloning and expression in E. coli of an exolevanase from
RT Acetobacter diazotrophicus SRT4."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; L41732; AAF16405.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 534 LEVANASE.
SQ SEQUENCE 534 AA; 58367 MW; 7BDBBA4A9BA478F6 CRC64;

Query Match 18.1%; Score 513; DB 2; Length 534;
```

```
Best Local Similarity 30.2%; Pred. No. 7e-26;
Matches 165; Conservative 75; Mismatches 238; Indels 68; Gaps 22;

Qy 5 ISRAVLGAGAGALIFGGAAPPAARASAPGSLRAVYHMTTPPSGWLCDPQRPVTHCA 64
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 16 LSRRCL-----PAAFLFLPLGAPHA--ADTP-QWRPVLFHSPAPAYWMDNPGPILLGV 67
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 65 YOLYYLHS--DQNGPGGDHASTTGDVAFTHHGTVMLPDPFPVWGSAAVGTANTAGF 122
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 68 YHLYFOYAPGSMWGHPSMGHATSTDLHWHTEHGVIAAATPGCEIFSGSLVPDLNRSGL 127
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 123 G---AGAVVALATQ-----PTDGVRYQOYLWSTGDGFTTALP-DPVI-VNTDGR 170
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 128 GSTDAPPLLAFTSVFHDNPAHPDGT---QAQSVSVSHDGGFTWRPYAHNPVLTILHPSR 184
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 171 AATTPAEIENAEWFRDPKIHWDTARGEVW--CVIGRLRYAAFYTPNLRDWTLRNFDY 228
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 185 Q-----FRDPSVFWYDGGCWMTTVVGDALVKLYRSTDLHWSFLSDFQPS 232
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 229 NHALGGI--ECPDLEITADGTR--HWVLAASMDAYGIGLPMYVYWTGWDGEOFHAD 284
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 233 GYRFGMLWEMPTLVPLKLDGNPRATRWVMTVSVNPNWSIAGGSGVQYFVGRFDGVTFTPD 292
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 285 DLTP-----QWLDWGDHAYAAVTPESI--DAPETKRLAIAWNNKYAARDVPTDAS 334
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 293 ALPPPGSDPSRYDWLDHGDAGQYATTLFANTGSGAP----VLIGWMDNDYAT-DLPTAP- 346
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 335 DYGNGQNSIVRELRLARQPGGWYLLSTPVAALTYV--TATTLPLDRTV--DGSVALPW 390
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 347 --WRGQWTLVDIAL-KTVDGHPVVIQTPTRAYEDMVRKKGWVYVYDGNLPODGRVTIPS 403
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 391 NGRAYEIELDIAMDTATNVGSGRSPDGTNRHTNIGKYGAD--LYVDRGSDLAGYSLAP 448
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 404 RGEVLDIRLVLRGARRAGVIRETPDGTGTSVSYDFVDGTLTVDRGSLGVGFSRPF 463
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 449 YSRAAAPIDPGARSVHLRLVDTOSVEFVNAGHTVLSQQVHFABEGDTCISLYTDGGRAH 508
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 464 STRHAYLAAPGGEVALHVVDRASVELFANDGVLRMTDLIFPPAGSDRISLFAEGGAT 523
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 509 FTGIWV 514
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 524 IHGLRV 529
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 12
Q8RLU2 PRELIMINARY; PRT; 485 AA.
ID Q8RLU2 AC Q8RLU2
DT 01-JUN-2002 (TremBrel. 21, Created)
DT 01-JUN-2002 (TremBrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBrel. 21, Last annotation update)
DE Exo-inulinase.
GN INU3.
OS Xanthomonas oryzae.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGL21;
RA Kim B.-W., You D.-J., Kwon H.-J.;
RT "Nucleotide sequence of exo-inulinase from Xanthomonas oryzae MGL21.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY077613; AAL82575.1;
SQ SEQUENCE 485 AA; 5525 MW; E0AD6E7F93E0187B CRC64;

Query Match 17.1%; Score 486.5; DB 2; Length 485;
Best Local Similarity 32.0%; Pred. No. 3.4e-24;
Matches 157; Conservative 74; Mismatches 201; Indels 59; Gaps 24;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQYLYLHSDGNN--GPGGDHASTTDCVATHTGHT 97
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 10 RPSVHFSPKGMNDNPGMWYVEGRYHLYQHHPFGTTGPMHWGHAVSTDLMNWEEQ-- 67
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
Qy 98 VMPLRPD--FPWWSGSAVVGTAANTAGF--GAGAVVALATQ-----PTDGVRYQOYLW 148
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 68 PIALEPDELGITFGSSAVVDEQDTSFGFGKPLVAIFTHHSLPGTEQVR--QCOSLAY 125
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 149 STDGGFTTALP-DPVI-VNTDGRAATTPAEIENAEWFRDPKIHWDTARGEVWCVIGRLRY 207
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 126 STDSKTKWIKYTGPNVL-----EDKECIDFRDPKVFWMHKPTQWVMVLACGQT 173
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 208 AAFYTPNLRDWTLRNFDYPNHALGG---IECPDLEITADGTRH--WVLAASM-DA 260
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 174 TRIYHSPNLKEMTFASEF--GHGISHDVAWVECPDLEPLYIDEREQVWVWMLVSGDV 230
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 261 YGIGLPMTYAYWTGWDGEOFHADDLTP--QWLDWGDHAYAAVTPWPSIDAPETKRLAIAW 318
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 231 PEIREGSRQYFTGFDGTTFFVADAASEKVRWLDYGRNYAGVCSWDIPAEDGRLEFMGW 290
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 319 MNNWYAADVPDTPDASDYGNGQNSIVRELRLARQPGGWYLLSTPVAALTYVYATTLPL 378
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 291 MSNWRV-ANQTPTR--WRGAMSIPELALETRK-GTVALVQRPVRELEGLRTPVLSLT 345
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 379 DRTVD--GSVLPWNGRAYEIELDIAMDTATNVGSGRSPDGTNRHTNIGKYGAD--DLV 434
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 346 EPSWEVNRNALSALQDCYELVAEFA--TTGDFGFKVRVSD--OETLVGYTSATQEVYV 401
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 435 DRGSPDLAGY--SLAPYSRAAPIDPGARSVHLRLVDTOSVEFVNAGHTVLSQQVHFA 492
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 402 DKRSGSRDFEDHPAGKHGASLTILPD--RMDIRIYVDRSSIEVFFDRGQVAITDLI-PP 458
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 493 EGDY-GISLYT 502
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 459 DAEAKGLEIFS 469
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 13
Q9APY5 PRELIMINARY; PRT; 493 AA.
ID Q9APY5 AC Q9APY5
DT 01-JUN-2001 (TremBrel. 17, Created)
DT 01-JUN-2001 (TremBrel. 17, Last sequence update)
DT 01-MAR-2002 (TremBrel. 20, Last annotation update)
DE Exo-inulinase (EC 3.2.1.80).
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SNU-7;
RA KOO B.-S., Kang S.-I., Kim S.-I.;
RT "Nucleotide sequencing of the gene for exo-inulinase from Bacillus sp.
snu-7 and its expression in Escherichia coli.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL: AF234992; AAK00768.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 493 AA; 53604 MW; 61E44A79A1600447 CRC64;

Query Match 16.5%; Score 469.5; DB 2; Length 493;
Best Local Similarity 29.4%; Pred. No. 4.6e-23;
Matches 148; Conservative 70; Mismatches 212; Indels 73; Gaps 18;

Qy 15 GAGALALIFGGAAPPAARASAPGSLRAVYHMTTPPSGWLCDPQRPVTHGAYQYLYHSDQ 74
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 17 GYGKLTSL--GTAASLAAAITVG--RPILHYTAKTNLWMDNGLVWGHGVHLCFTQNNPF 72
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 75 NN--CPGQWDHASTTGDVAFTHHGTVMLPDPFPVWWSGSAVVGTAANTAGFAGAGAVVALAT 132
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 73 DNVGNNWGHATSTDLHWHTEHGPVIAACNEEDVFSSIVVDHGTSGFTAEADPALVA 132
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 133 QPTDGVK-----YQEQYLWSTDGGFTTALPDPIVNTDGRAATTPAEIENAEWFRD 186
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```



|           |        |   |              |
|-----------|--------|---|--------------|
| Db        | 181    | PKVFRYEGPAGSRWITWAVEAQHOQVVLVRSADLKDWEYLSITGCPANASGEWECBP--LP | 238          |
| Qy        | 243    | ITADGTRHWWLAASMDAYGICLPMTYAYTGTWIDGQGFHA--DOLTPQ-----         | 289          |
| Db        | 239    | VDGDPDNKWWLVNINPGAVAGSGGQYFVGDFGVQFPADPDSLVPPTDAGNDVLRHC      | 298          |
| Qy        | 290    | -WLDGWGDWYAAVWTWPSIDAPETRKLATAWNNNKYAARDVPTDASDGYNGQNSIVRELR  | 348          |
| Db        | 299    | LWLDGWRDYAAVSFS--NAPENRRIMIGMNNWDY-ANSLPTSP--WRSSMSLAREIE     | 352          |
| Qy        | 349    | LARQGGGWYTLTSTPVAALTNYVTATT-----TLPDRTVGGSVAVLPWNGRAYEI       | 397          |
| Db        | 353    | LA-TVDGFPRLVQRPVLPLDSDGEPARTIQNMELTTPCCNCPTRRPPGSAQL-----I    | 403          |
| Qy        | 398    | ELDIAMDATNNGISVGRSPDGTGTRHT--NIGKYGADLYVDRGPSDLAGYSLAPYSRAAP  | 455          |
| Db        | 404    | EAEILPGTARHIAFRLGAPDGSAAATVLSFDALTSRLTLDRRNSGNTAFTISSRLNPQP   | 463          |
| Qy        | 456    | IDPG-ARSVHLRILVDQGSVEVF 477                                   |              |
| Db        | 464    | VKASTGGALRLKVIIDQGSVEVF 486                                   |              |
| RESULT 15 |        |   |              |
| ID        | O44109 | PRELIMINARY:  | PRT: 943 AA. |

```

DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Levansase.
GN LEVJ.
OS Actinomyces naeslundii.
OC Bacteriia; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TI4V;
RX MEDLINE=951129923; PubMed=7828936;
RA Norman J.M., Bunny K.L., Giffard P.M.;
RT "Characterization of levJ, a sucrose/fructanase-encoding gene from
RT Actinomyces naeslundii TI4V, and comparison of its product with other
RT sucrose-cleaving enzymes.";
RL Gene 152:93-98(1995).
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; U12274; AAA67876.1; -.
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 943 AA; 93267 MW; 801A2A7D074579B9 CRC64;
Query Match 16.5%; Score 467.5; DB 2; Length 943;

```

|    |                       |  |                    |                                      |
|----|-----------------------|--|--------------------|--------------------------------------|
|    | Best Local Similarity | 21.9%;   | Pred. NO. 1.5e-22; |                                      |
|    | Matches               | 187;   | Conservative       | 83; Mismatches 217; Indels 381; Gaps |
| QY | 4                     | AISRAVLQGAGACALALIFGG-----APVPAARASAFGS-LRAYVIHMTPPSGWLCDPQP 58      |                    |                                      |
| DB | 19                    | AASTLASMPAAASGQARPVSAAPNAPAPTAPKADOTGERWRPOSHYTPQKNMNDPNGL 78        |                    |                                      |
| QY | 59                    | VTHGAYQLYLHSDDNN--GPGGDWHAISTDGVAFTHHGTVMLRPDPFVWSGSVAVGTT 116       |                    |                                      |
| DB | 79                    | VYYDGEYHMFYQNPEGSDWGNMGHWASHAVSKDLVHQELGVALPHYSQYCVFGSSAVIDT 138     |                    |                                      |
| QY | 117                   | ANTAGFGA--GAVVALATQPTDCVRKYQRQYLWSTDGDTFTAL--PDPIVTNTDGRA 171        |                    |                                      |
| DB | 139                   | KNTSLGSGSPNPAMVAVMTRADVCGN--QSOSLAYSTDKGTWNLYNNGDPVL----- 189        |                    |                                      |
| QY | 172                   | ATTAEATENAEWFDRPDKIHMDTARGEWCVCIGRL--RYAAFYIYTNLRTWTLRNFEDYN 229     |                    |                                      |
| DB | 190                   | ----DIGSNE-FRDPKVFVDQAQSGRWITMVVSHATEHRVSYFSYSPDLLIIHWTEOSSFGGEG 243 |                    |                                      |



|    |     |  |            |     |
|----|-----|--|------------|-----|
| QY | 230 | HALGTECEPDLFEITADGCTRH--WVLA-                                  | -----      | 255 |
|    |     | : : :  |            |     |
| Db | 244 | ITSAVWACPDFPLPVDGSSQVKWLVVTVADSAQYFVGSWDGTTFTPDEIPHYSBEG       | 303        |     |
|    |     | -----  |            |     |
| QY | 256 | -----ASMDAYG-  | -----      | 262 |
|    |     | :  :   |            |     |
| Db | 304 | TTLADFENCYAGWKADCAAFGSPATGDLPGHGKAYVDSFGSGDADTGTLTSDFTVSS      | 363        |     |
|    |     | -----  |            |     |
| QY | 263 | -----  | -----      | 262 |
|    |     | -----  |            |     |
| Db | 364 | SYINLRTAGGKHPYNPQATGNDGGKRLLAGFDGSEGWTVEGSAFAATPPQCATPAQOPL    | 423        |     |
|    |     | -----  |            |     |
| QY | 263 | -----IGLPWTY-  | -----AY-   | 271 |
|    |     |  |            |     |
| Db | 424 | VHSSAGLNTNLTDAATGOGSDAPTGTATPTFTTIDSAYLNLLMGNNPRPEGGADGGS      | 483        |     |
|    |     | -----  |            |     |
| QY | 272 | -----W-  | -----TGTW- | 276 |
|    |     |  |            |     |
| Db | 484 | RVSVELIYVDGKVRSATGRNLEELNQSWDYSDLKGSQIIVTDTATGGWGHLLDDEV       | 543        |     |
|    |     | -----  |            |     |
| QY | 277 | -----  | -----      | 276 |
|    |     | -----  |            |     |
| Db | 544 | RASDKKASPIADNTSVNLVVDGKVVASATGNNSGTLEWTSMKVAAKGRKARLVIEDRNG    | 603        |     |
|    |     | -----  |            |     |
| QY | 277 | -----DGEQFHADDLTPQWLDGWDMWAAVTPSIDAPETKRLAIAMNN                | 320        |     |
|    |     | -----  |            |     |
| Db | 604 | NAEDGCHLMVDQILQSDTKAFSGADVPR-LDYKDYAAVTDWNV--PNCKRRYQVGWMS     | 660        |     |
|    |     | -----  |            |     |
| QY | 321 | NKYYAARDVPTQASDGYNGONSIVRELIAROPGGWYTLTSTPVAALNYVTA-----T      | 374        |     |
|    |     | : : :  |            |     |
| Db | 661 | NWAY-VRDLPTTT--WRTAMSTVREMGLTR-VNGKRLTAQVPTALESLRTGOELIRKD     | 715        |     |
|    |     | -----  |            |     |
| QY | 375 | TTLP-LDRTVDGSAVLPNWNGRAVEIFELDTAWDTATNVGLSVGSRPDGTRHNLG--KYCAD | 431        |     |
|    |     | : : :  | :  :       |     |
| Db | 716 | TDIPVGETSLGRA--OQTSLDISVDSPSSAFGLKV--LDNGEQYTLIGYDSQAKQ        | 770        |     |
|    |     | -----  |            |     |
| QY | 432 | LYVDRGPSDLAGYSLAPYSRAAAPIDPGAR-SVHLRLIIVDTQSVFVFNAGHTVLSQOVH   | 490        |     |
|    |     | : : :  | :  :       |     |
| Db | 771 | LVVDRTHSGVTDSPKPEPARSTAPLSPSKGOVHLRIIIVDAHSVEVFAADGTPVITQIV    | 830        |     |
|    |     | -----  |            |     |
| QY | 491 | FAEGDTGISLYTDGGPAHFTGIVVREIG                                   | 518        |     |
|    |     | :  :          :  :   |            |     |
| Db | 831 | PQODATGYSLYAEGGTAHLGSLSLMHLG                                   | 858        |     |

Search completed: May 6, 2003, 12:34:43  
Job time : 58.8904 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:31:02 ; Search time 19.4625 seconds  
(without alignments)  
784.611 Million cell updates/sec

Title: US-09-868-328B-4  
Perfect score: 2837  
Sequence: 1 MTPAIRRAVLQAGAGALA.....LYTDGGAHPTGIVVREIGQ 519

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 499.5 | 17.6        | 943    | 4     | US-09-397-885-5     |
| 2          | 468.5 | 16.5        | 1277   | 4     | US-09-397-885-3     |
| 3          | 461   | 16.2        | 923    | 4     | US-09-397-885-1     |
| 4          | 348   | 12.3        | 581    | 4     | US-09-331-581-2     |
| 5          | 333.5 | 11.8        | 337    | 4     | US-09-331-581-24    |
| 6          | 291   | 10.3        | 332    | 4     | US-09-331-581-23    |
| 7          | 275   | 9.7         | 635    | 1     | US-08-245-809-1     |
| 8          | 274.5 | 9.7         | 635    | 1     | US-08-296-624-2     |
| 9          | 235.5 | 8.3         | 492    | 4     | US-09-134-001C-3895 |
| 10         | 205.5 | 7.2         | 630    | 3     | US-08-860-091A-2    |
| 11         | 205.5 | 7.2         | 630    | 4     | US-09-245-323A-6    |
| 12         | 189   | 6.7         | 615    | 3     | US-08-860-091A-4    |
| 13         | 189   | 6.7         | 615    | 4     | US-09-245-323A-8    |
| 14         | 166.5 | 5.9         | 626    | 4     | US-09-019-385-2     |
| 15         | 147.5 | 5.2         | 565    | 4     | US-09-142-623-11    |
| 16         | 141.5 | 5.0         | 635    | 4     | US-09-142-623-1     |
| 17         | 140.5 | 5.0         | 574    | 4     | US-09-142-623-13    |
| 18         | 119.5 | 4.2         | 3739   | 3     | US-09-320-878-2     |
| 19         | 117   | 4.1         | 3491   | 2     | US-07-642-734C-2    |
| 20         | 117   | 4.1         | 3491   | 3     | US-08-439-009A-2    |
| 21         | 115   | 4.1         | 3739   | 4     | US-09-105-537-33    |
| 22         | 115   | 4.1         | 11877  | 4     | US-09-105-537-6     |
| 23         | 110.5 | 3.9         | 983    | 1     | US-08-167-919A-10   |
| 24         | 110.5 | 3.9         | 983    | 2     | US-08-449-645A-21   |
| 25         | 110.5 | 3.9         | 983    | 2     | US-08-702-367A-21   |
| 26         | 110.5 | 3.9         | 983    | 3     | US-08-715-106-10    |
| 27         | 110.5 | 3.9         | 983    | 5     | PCT-US95-04681-21   |

|    |       |     |      |   |                   |                    |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 106   | 3.7 | 3519 | 4 | US-09-428-517-4   | Sequence 4, Appli  |
| 29 | 106   | 3.7 | 4551 | 3 | US-09-320-878-1   | Sequence 1, Appli  |
| 30 | 106   | 3.7 | 4613 | 4 | US-09-105-537-31  | Sequence 31, Appli |
| 31 | 105.5 | 3.7 | 811  | 1 | US-08-480-604A-7  | Sequence 7, Appli  |
| 32 | 105.5 | 3.7 | 811  | 2 | US-08-405-496A-7  | Sequence 7, Appli  |
| 33 | 105.5 | 3.7 | 811  | 4 | US-08-915-136-7   | Sequence 7, Appli  |
| 34 | 105.5 | 3.7 | 811  | 4 | US-08-957-310-7   | Sequence 7, Appli  |
| 35 | 105.5 | 3.7 | 812  | 1 | US-08-480-604A-29 | Sequence 29, Appli |
| 36 | 105.5 | 3.7 | 812  | 4 | US-08-915-136-29  | Sequence 29, Appli |
| 37 | 105.5 | 3.7 | 2710 | 1 | US-08-480-604A-6  | Sequence 6, Appli  |
| 38 | 105.5 | 3.7 | 2710 | 2 | US-08-405-496A-6  | Sequence 6, Appli  |
| 39 | 105.5 | 3.7 | 2710 | 4 | US-08-915-136-6   | Sequence 6, Appli  |
| 40 | 105.5 | 3.7 | 2710 | 4 | US-08-957-310-6   | Sequence 6, Appli  |
| 41 | 104.5 | 3.7 | 964  | 4 | US-08-484-791-2   | Sequence 2, Appli  |
| 42 | 104   | 3.7 | 655  | 4 | US-09-311-626B-14 | Sequence 14, Appli |
| 43 | 104   | 3.7 | 3033 | 1 | US-07-925-695-5   | Sequence 5, Appli  |
| 44 | 103   | 3.6 | 936  | 5 | PCT-US94-05905-22 | Sequence 22, Appli |
| 45 | 102.5 | 3.6 | 614  | 1 | US-08-262-338A-4  | Sequence 4, Appli  |

ALIGNMENTS

RESULT 1  
US-09-397-885-5  
; Sequence 5, Application US/09397885  
; Patent No. 6323007  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rahbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And  
; TITLE OF INVENTION: Process For Using The Enzyme  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/397.885  
; CURRENT FILING DATE: 1999-09-17  
; EARLIER APPLICATION NUMBER: PA 1998 01173  
; EARLIER FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: PA 1998 01623  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 60/101.615  
; EARLIER FILING DATE: 1998-09-24  
; EARLIER APPLICATION NUMBER: 60/111.675  
; EARLIER FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-09-397-885-5

Query Match 17.6%; Score 499.5; DB 4; Length 943;  
Best Local Similarity 30.8%; Pred. No. 9.1e-37;  
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;

|    |     |   |     |
|----|-----|---|-----|
| QY | 40  | RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNGPGGWDHASTTDGVAFTHHGTVM   | 99  |
| DB | 383 | RPDYHSPARGSADPNGLVVEYGEYHLF--HQD-----GGTWAHAVSTDLVHWKRLPTAL | 436 |
| QY | 100 | PLRPDPFPPVMSAVVGTANTAGF-----GAGAVVALATQPTDGVKRYEQVLYWSTDGGF | 154 |
| DB | 437 | PWNDLGHVWSGSAVDLHNASGLFADSGKGLIAYTTSYNDPRPNQRIGLAYSKDRGR    | 496 |
| QY | 155 | TFT-ALPDPIVINTDGRAATTPAEIENAEW-FRDPKIHMDTARGWCVIGRLRYAAYT   | 212 |
| DB | 497 | TWEYAAERPIVIENPKQGGDP-----GGWDFRDPKVVDEHNRWVWVSGGDHIREFT    | 551 |
| QY | 213 | SPNLRDWTLRRNFDYPNALGGI-ECPLFEITADD-GTRHWVLAASMDAYGIGLPMYTA  | 270 |
| DB | 552 | STNLIDWTLTDSFGYGVYRGVWECPLQLFQLAVDDTGEKKVLMISTGAN-----PNTQG | 607 |

```
QY 271 YWTGWDGEQFHADLLTPQW-----LDWGWDWYAAVTPSIDAPETKRLAIA 317
      : : : : :
Db 608 -----SAAEYFIGELTPECKFVNDNPAGKVLATDYGKEYYASMSFAGM--PDGRRVMLA 659
QY 318 WMNNKVAARDVPTDASDGYNGQNSIVRELRLARQPGCWYTLTSTPVAALTN-----YVTA 373
      : : : : :
Db 660 WMNNWYDFFA-FPT---EGWKGLVSLPRELTQKTDKG-IRLAQTPIRELESIRGOLLFA 714
QY 374 TTTLPDRTVDG---SAVLPWNGRAYEIELDIAMDATNNGISVGRSPDGRHNTICKYGA 430
      : : : : :
Db 715 AS---DRVQADRENLLKGSSGYEIEAIEIPQASNSWSEFGRRLREGAGKRTVVGKYT 771
QY 431 ---DLVYDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSVFVNAGHTVLSQ 487
      : : : : :
Db 772 KENEIYVDRSLSGDTGFSERFTTLHOAPLOPDNRNRKRLIRFVDSLSLEVEGGGRVFESE 831
QY 488 QVHFAEGDTGISLYTDGGAHFTGIIVREI 517
      : : : : :
Db 832 VIFPDPAHREMSLFTVGGEVNVVSLKVHAL 861

RESULT 2
US-09-397-885-3
; Sequence 3, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; EARLIER FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Paenibacillus pabuli
US-09-397-885-3

Query Match 16.5%; Score 468.5; DB 4; Length 1277;
Best Local Similarity 29.1%; Pred. No. 9.5e-34;
Matches 168; Conservative 74; Mismatches 243; Indels 93; Gaps 25;

QY 4 ATSRRAVLQAGAG-----GALIFGAGVAPPA-----RASAPG----- 37
      : : : : :
Db 320 ASSNRKVPSPGARRHIEVIASSLLQVYVDGTPAAVEVTDKSYAKGNAGLVQOGMAY 379
QY 38 -----SLRAYHMTTPSGMCLDQRPVTHGAYQLYLLHSDQNGGWDHSTTDCGVAFTHHGTVM 99
      : : : : :
Db 380 FQDIYMTESMYKENYRQYHYSPLRGSASDPNGLVYFEGEYHLP---HQD-----GGTWA 433
QY 83 HASTTDGVAFTHGTVMLRDPFVWSSGSAVVGCTANTAGF-----CAGAVVALATQPTDG 154
      : : : : :
Db 434 HAVSSDLINWKLPIALPNDGHWSSGSAIADLNNSGLFTSDSGKGLIAYTSTVHPDK 493
QY 138 VRKYQBYLWSTDGGFTFT-ALPDPVIVNTDGRAATTPAEIENAEW-FRDPKIHWDTAR 195
      : : : : :
Db 494 PGNQIRIGLAYSTDQGRNQYAKERIVIDNPKNGDDP-----GSWDFRDPKVVREDEH 548
```

```
QY 196 GEWCVCIGRLRYAAFTYSPNLRDWTLLRRNDYFNHALGGI-ECPDLFETAD-DGTRHWV 253
      : : : : :
Db 549 NRWVVVSGDGHIRFETSTNLLDWTILTNFYGDYVRGGWECPDLIQLPVDGTGQRKW 608
QY 254 LAASMDAYGIGLPMYTA-----YWTG--TWDEQFHADLLTPQWL--DWGHWYAAVTPWS 305
      : : : : :
Db 609 LLISTGAN-----PKTQGSDAEYEVGQLTADG-KFLNDHPAGQVLRDYGKEFYASMSFAN 663
QY 306 IDAPETKRLAIAWMNNKVAARDVPTDASDGYNGQNSIVRELRLARQPGCWYTLTSTPVA 365
      : : : : :
Db 664 M--PNQRKVMYLAWMNTWYDFF-EFTSS---WKQGLTIPREVSRLTDEG-VRLVQTPIT 716
QY 366 ALTNVYATTTLPDRTVDGSAVLPWNG---RAYEIELDI---AMDATNNGISVGRSPDG 419
      : : : : :
Db 717 ELQKLRLHNSYSAQOMTVGPKSNPLEGLTAGAYEIEAEVEIPANSSVTEFGQL-RQREG 775
QY 420 TRITNICKYCAD---LYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVHLRLIVDTQSV 476
      : : : : :
Db 776 QKTT--VAYRVDTONMFVDRTTSGDVSFSDLETKVHEASLKPENQKVKLRFVDESSEV 833
QY 477 FVNAGHTVLSQQVHFAEGDTGISLYTDGGAHFTGIWV 514
      : : : : :
Db 834 FGNDGKVFSDVIFPDPAGRAMAFYSLGGEVKVSSMKV 871

RESULT 3
US-09-397-885-1
; Sequence 1, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; EARLIER FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 923
; TYPE: PRT
; ORGANISM: Paenibacillus amylolyticus
US-09-397-885-1

Query Match 16.2%; Score 461; DB 4; Length 923;
Best Local Similarity 29.5%; Pred. No. 2.8e-33;
Matches 150; Conservative 74; Mismatches 223; Indels 62; Gaps 23;

QY 40 RAVYHMTTPSGMCLDQRPVTHGAYQLYLLHSDQNGGWDHSTTDCGVAFTHHGTVM 99
      : : : : :
Db 392 RPQHYHTPIRGASDPNGLVYFEGEYHLP---HQD-----GGTWAHVAVSKDMLNKRPIAL 445
QY 100 PLRDPFVWSSGSAVVGCTANTAGF-----CAGAVVALATQPTDGVKRYQBYLWSTDGGF 154
      : : : : :
Db 446 PNDHGHVWSSGSAVADMTNASGLFGDSGKGLIAYTSTFNPSPNGNQIRIGLAYSKDQGR 505
QY 155 TPTALPD-PVIVNTDGRAATTPAEIENAEW-FRDPKIHWDTARGEVCVIGRLRYAAFTY 212
      : : : : :
Db 506 TWEYSKERPIVENPKSGN-----EAGNDFRDPKVRIRDDNNRWMVWVSGDGHIRFET 560
QY 213 SPNLRDWTLLRRNDYFNHALGGI-ECPDLFETADDT--RHWVLAASMDAYGIGLPMY 268
```

Db 561 STNLLDWTLTDMWGYGVGWECPDLFQLPV--DGTSSQKWMVMIETGAN----PKTG 615  
Qy 269 ---YAYWTG--TWDEQFHADDLTPOWL--DWGMDWYAAVTPSIDAPETKRLAIWANN 321  
Db 616 GSDAEYFIGHLTADG-KFVNDNPAGKVLRTDEGREFYASMSFANM--PDHRTVMWMTN 672  
Qy 322 WKYAADVPTDASDGYNGNSIVRELRLAROPGGWYTLTSPVAALTYVATTPLPRT 381  
Db 673 WDYPPA-PPT---SNWKGELTIPREVSIVTTEDG-IRMVQSPIKEL-----ESLRKPLYS 722  
Qy 382 VDGSAVLPNGR-----AYEELDI---AWDTATNVGIVSGRSPDGRHNLGKYGA 430  
Db 723 ASNKSVPSSNGLKGIISGAYEIAEIEIPETSTVTEGFEINRG--ANKQTVVGYKAS 780  
Qy 431 D--LYVDGSPDLAGYSLAPYSRAAAPIDPGARSVHLRLVDTQSVVEFVNAAGHTVLSQ 488  
Db 781 DSRMFVDTASGETDFSNLFSKKHEAPTQMNENRILKMLIVDESSVEAFNGDKGVVFSOV 840  
Qy 489 VHFAGDGTGSLYTDGGPAHFTGIVVREI 517  
Db 841 IFPPASRAMSFYKGVGNVNVSLKVHQL 869  
  
RESULT 4  
US-09-331-581-2  
; Sequence 2, Application US/09331581  
; Patent No. 6130070  
; GENERAL INFORMATION:  
; APPLICANT: TOHDA, Hideki  
; APPLICANT: HAMA, Yuko  
; APPLICANT: KUMAGAI, Hiromichi  
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE  
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS  
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF  
; FILE REFERENCE: 0059-1142-0PCT  
; CURRENT APPLICATION NUMBER: US/09/331.581  
; CURRENT FILING DATE: 1999-06-30  
; EARLIER APPLICATION NUMBER: PCT/JP98/04929  
; EARLIER FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: JP 9-314608  
; EARLIER FILING DATE: 1997-10-31  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 581  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
US-09-331-581-2  
  
Query Match 12.3%; Score 348; DB 4; Length 581;  
Best Local Similarity 24.9%; Pred. No. 2.7e-23;  
Matches 132; Conservative 82; Mismatches 226; Indels 90; Gaps 22;  
  
Qy 28 PPARASAPGSU-----RAYHMTTPSGMCLDPPQRPVTHGAYQLY 69  
Db 53 PPFVNTTAPNCTCLGNEXYLPSPGYNATDRPKIHFTPSSGFMNDPGLVYTGYYHMF 112  
Qy 70 LHSQD--NNGPGWDHASTTDCVAF-----THGTVMLRPLRDPFVWGSAAVGTGA 117  
Db 113 QYSPKTLTAGEVHMGHTYVKDLIHENYPIAIPDEHENGVLSL-----PFGSAAVVDVH 167  
Qy 118 NTAGFGAG-----AVVALATQPTDGVKRYOYLWSTDGGFTTALPDPVIVNTDGRA 171  
Db 168 NSSGLFSNDTTPERIVLIYDHTWTGVA--ERQAIAYITDGGYTK-----KY 213  
Qy 172 ATTPAEIENAEWFRDPKIHWDTARGEWCYVIGRLY--AAFTYSPNLRLDWTLLRRNFDYFN 229  
Db 214 SGNPVLIDNSLOFRDPKPVINDFANRWVIVAMSONYGIAYSSVDLIHWELSVFSTSG 273  
Qy 230 HALGIECPDLFEITADGTRH--WYLAASMDAYGTLPM---TYAYWTGTWDEGOFHADD 285  
Db 274 YLGQYECFPGMARVPVEGTDEYKWLFTSINP---GAPLGGSVQYFVGVGDNWNGTNEVPD 330

Qy 286 LTPOWLDMGWDMYAAVTPSIDAPETKRLAIWANNMKYAAARDVPTDASDGYNGNSIVR 345  
Db 331 GQTRVLDLKGDFYASALYHSSA--NADVIGVGWASNWQY--TNAAPTQV---FRSANTVAR 385  
Qy 346 ELRLARQGGWYT---LLSTP--VAALTN--YVATTTLTDPRTVYDGSVAVLPWNGRAYE- 396  
Db 386 KFTLRDVPONPMTNLTSLIQTPLNVSLLRDETLFTTAPVINSSSLSGSPITLPSNTAFEF 445  
Qy 397 -IELDIAMDATNTNIGSVGR---SPDQTR---HTNIGKYGADLYVDRGPSDLAGY--SL 446  
Db 446 NVTLISINTEGCTGYCLGRILIIDSDPYRLQSIISVDVDFAASTLVINRAQMGWNSL 505  
Qy 447 APYSRAAAPIDPGARSVHLRLVDTQSVVEFVNAAGHTVLSQOVHPAEGDT 496  
Db 506 FTFSPANDIYIG--NVTLYGIVDNGLLLEYVNGEKTYTNDFFFLQAT 553  
  
RESULT 5  
US-09-331-581-24  
; Sequence 24, Application US/09331581  
; Patent No. 6130070  
; GENERAL INFORMATION:  
; APPLICANT: TOHDA, Hideki  
; APPLICANT: HAMA, Yuko  
; APPLICANT: KUMAGAI, Hiromichi  
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE  
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS  
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF  
; FILE REFERENCE: 0059-1142-0PCT  
; CURRENT APPLICATION NUMBER: US/09/331.581  
; CURRENT FILING DATE: 1999-06-30  
; EARLIER APPLICATION NUMBER: PCT/JP98/04929  
; EARLIER FILING DATE: 1998-10-30  
; EARLIER APPLICATION NUMBER: JP 9-314608  
; EARLIER FILING DATE: 1997-10-31  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-331-581-24  
  
Query Match 11.8%; Score 333.5; DB 4; Length 337;  
Best Local Similarity 30.0%; Pred. No. 2.5e-22;  
Matches 105; Conservative 55; Mismatches 139; Indels 51; Gaps 16;  
  
Qy 24 GGAVPPAARASAPGSLRAYHMTTPSGMCLDPPQRPVTHGAYQLYVHSDQNG 77  
Db 10 GFAAKMSASMTNETSDRPLVHFTPNKGMNDPGLWYDEKDAKWHYFQ--YNPNDTVMG 67  
Qy 78 -PGGDHASTTDCVAFTHHG--TVMLRPLRDPFVWGSAAVGTANTAGFCAGAV-----VA 129  
Db 68 TPLFWGHATSDDLTWNEDQPIAIPKRNDSGAFSGSMVVDYNNNTSGFFNDITDPQRCVA 127  
Qy 130 LATQPTDGVKRYOYLWSTDGGFTTAL--PDPVIVNTDGRAATTPAEIENAEWFRDPK 188  
Db 128 IWTYNTF---ESEEQYISYSTDGGYTFTEYQKNPVA-----ANSTQFRDPK 171  
Qy 189 IHWDTARGEWCYVIGRLR--YAAFTYSPNLRLDWTLLRRNFDYFNHALGG--IDCPOLFEIT 244  
Db 172 VWYEPSPQWIMTAAKSQDYKIEIYSSDDLKSWKTESAF--ANEGLGYQVECPGLIEVP 229  
Qy 245 A--DDGTRHVLAAASMDAYGTLPMTYA---YWTGTWDEGOFHADDLIPOWLDMGWDMY 299  
Db 230 TEQDPSKSYWVWFISINP---GAPAGGSFNOYFVGSFNGTHFEAFDQNSRVVDFGKDYA 286  
Qy 300 AVTWPSIDAPETKRLAIWANNMKYAAARDVPTDASDGYNGNSIVRELRL 349  
Db 287 LQTFENTDPTYGSLALGIASWNEYS-A-FVPTNP---WRSSMSLVKRFSL 332



RESULT 8  
US-08-296-624-2  
; Sequence 2, Application US/08296624  
; Patent No. 5658773  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Klann, Ellen  
; TITLE OF INVENTION: Tomato Acid Invertase Gene  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Weber  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,624  
; FILING DATE: 26-AUG-1994  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/770,970  
; FILING DATE: 07-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A.  
; REGISTRATION NUMBER: 31,677  
; REFERENCE/DOCKET NUMBER: 2307E-036510US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-543-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 636 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-296-624-2

Query Match 9.7%; Score 274.5; DB 1; Length 636;  
Best Local Similarity 24.1%; Pred. No. 1.6e-16;  
Matches 137; Conservative 76; Mismatches 236; Indels 119; Gaps 28;  
Qy 2 TPAISRAVLQAGAGALALIFGAVPPAARASAPGS-LRAVYHMTTPSGWLCDPQRPVT 60  
Db 65 SPAPPSRGVSGVSKTFRDVAGASHVSYANSLMSQRTAYHFQPKNMNDPNGLPLY 124  
Qy 61 THGAYQLYLHSDQNN--GPGGDHASTTDGVAFTHHGTVP--LRPD-----FPVWSGS 111  
Db 125 HKGWTHLEYQYNPSAINGNTWGHVSKDLT---HWLYLPFAMVPDQWYDINGVWTGS 180  
Qy 112 AVVGTTANTAGFCAGAVVALATQPTDGVKRYOXYLWSTDGGFTTALPDPIV---VNTD 168  
Db 181 ATI-----LPDQGMILYTGDDTDY--VQVQNLAYPAN-----LSDPLLLDWNKFK 224  
Qy 169 GRAATTPTAEIENAEWFRPKIHW-DTARGEWCV-----IGRLRYAAFTSPNLRDWTURR 223  
Db 225 GNPVLVPPGGVGRDFRPTTAWTGPQNGWLLTIGSKIGTGVALVYTSNFTSFKL-- 282  
Qy 224 NFDYFNHALGFI---ECPDLFEITA-----DGTGRHWVLAASM-----DAYGIG 264  
Db 283 -LDGVLHVAVPTGMWECVDFPVSTKKTNGLTSTNGPGVKH-VLKASLDDNKQDHYAIG 340  
Qy 265 LPMYVAYWTGDCQEQFHADDLTPQWLDGWDWYAAVTPSIDAPETKRLAIWNNWKY 324  
Db 341 ---TYDLGKNKWTDPNPELDCGIGLRLDYG-KYYASKTF--YDPKKRRVLWGWIGTDS 394

Qy 325 AARDVPTDASDGYNGQNSIVRELRLAROPGGWYTWLLSTPVAALTNVYVATTTLPTDVTD- 383  
Db 395 ESADL-----QKGWASVQSIPTVLYDKKGTG--THLLQWPFVEIESLRVGDPTV--KQVDL 446  
Qy 384 --GSAVLPWNGRAYEIELDIADWT-----ATNNGISVGRSPDGTGRHTNICKYG-- 429  
Db 447 QPGSTIELLRVSAEELDEASFEDVKVALQGIIEADHVFSCSTGGGAASRGILGPGFY 506  
Qy 430 --AD-----LYVDRG-----PSDLAGYSLAPYSRAAAPIDPGARS 462  
Db 507 VIADQTLSELTPVYFYISKGADGRAETHFCADOTRSSEAPGVGKQVYSSVPVLDERHS 566  
Qy 463 VHLRILVDTQSVVEFVNAGHTVLSQQVH 490  
Db 567 --MRLLVDSHIVESPAQGGRTVITSRIY 592  
RESULT 9  
US-09-134-001C-3895  
; Sequence 3895, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134.001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3895  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3895

Query Match 8.3%; Score 235.5; DB 4; Length 492;  
Best Local Similarity 23.5%; Pred. No. 3.8e-13;  
Matches 123; Conservative 69; Mismatches 202; Indels 129; Gaps 24;  
Qy 40 RAVYHMTTPSGWLCDPQRPVTHGAYQLYLHS-----DQNNPGPGGDHASTTDGVAFTHH 95  
Db 35 RQTHIQPEMGLLNDPNGLIYNGHY--YISHOWFPLGAVHGLKYWFYKSKDLLHFEPO 92  
Qy 96 GTVMPLRPD-----FPVWSGSVAVGTANTAGFCAGAVVALATQPTDGVKRYOEQ--YLYW 148  
Db 93 GTL--LKPDTRYDSHGVSYSASF-----EYQNHLYMYVT 124  
Qy 149 STDGGFTTALPDPIV--VNTDGRAATTPTAEIENAE-----WPRDPKTHWDTARGEWCV 200  
Db 125 GNRDQHNRISSOMIARNKNDGKTEKPKPVTHGQPGYTSHFDPKVF--EKNSOLYA 182  
Qy 201 VIG-----RLRYAAYTSPNLRDWTLRNFDPYFNHALGGI-----ECPDLFEIT 244  
Db 183 ILGAQENEMGRLLLYRSQDVVDW-----HFEGEIKTNLTQFGYMWECPDYFRLS 232  
Qy 245 ADDGTRHWVLAASMDAYGIGLPMY--AYWTGTDGEGFHADDLTPQWLDGWDWYAAVT 302  
Db 233 NKDVI--LMCPQGVAEAGDKFRNIYQSGYMGIDLNFNNLFFDHESFQELDNGDFEYAPOT 290  
Qy 303 WPSIDAPETKRLAIWNNWKYAADVDVTDASDGYNGQNSIVRELRLARQPGWYTLST 362  
Db 291 F--VDA-DGQRLIGWVG---LPDTEPTD-KEGWAHCLTIPRVLTENG-----KLKR 338  
Qy 363 PVAAL---TNVVTATTTLPDRTVDGSA-----VLPWNGRAYEIELEIDAMDTATNVGIS 412  
Db 339 PFQLEDLRTNKETAL-----GYANKPKRLHPYEGKQYEMIIDILENDA-EIYFE 389  
Qy 413 VGRSPDGTGRHTNICKYGADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSV-----HL 465

```

Db      390 LRSSRSESLITVKNHKNKLTDRDTS-----GTLPSNVGDGTTTRSTILDSPKLQOL 433

QY      466 RILVDTSQSVFVNAGHTVLSQQVHPAEGDTGISLTYTDGGPAH 508
        :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
Db      440 QIFVDTSIEIFCNDGERVLTSRIFFPNEDATGIKASTESGVY 482

RESULT 10
US-08-860-091A-2
; Sequence 2, Application US/08860091A
; Patent No. 6057494
; GENERAL INFORMATION:
; APPLICANT: Koops, Dr. Andries Jurriaan
; APPLICANT: Van der meer, Dr. Ingrid Maria
; APPLICANT: Van Tunen, Dr. Arjen Johannes
; TITLE OF INVENTION: DNA sequences encoding carbohydrate
; polymer synthesizing enzymes and method for prod
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LADAS & PARRY
; STREET: 26 WEST 61 STREET
; CITY: NEW YORK
; STATE: NY
; ZIP: 10023
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/4" Floppy disk 1.44MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Microsoft Windows for Workgroups
; SOFTWARE: Wordperfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,091A
; FILING DATE: 02-JULY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL/00012
; FILING DATE: 08-JANUARY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Cord, Janet I.
; REGISTRATION NUMBER: 33,778
; REFERENCE/DOCKET NUMBER: U-011329-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 708-1935
; TELEFAX: (212) 246-8959
; TELEX: 233288
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-860-091A-2

Query Match 7.2%; Score 205.5; DB 3; Length 630;
Best Local Similarity 21.2%; Pred. No. 2.9e-10;
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps

QY      20 ALIFGAVPPAARASAPGSL-----RAYVHMTTPSGWLCDPQPVTTHGAYQLYYL 70
Db      65 ATTFVTQLNIDLKRVPGLKDSAEVEWQRSTYHFQPKNFISDDPGPMYHWGTHLFYQ 124
QY      71 HSDNN--GPGGWDHASTTGDVAFTHHGTVMP--LRPDFPVMSGSAVVGTANTAGFGAGA 126
Db      125 YNPQSAIMGNITGWHSVKDMINWFH---LPFAMVPDH--WYDIGEVMTCGSATVLPNGQ 178
QY      127 VALATQPTDGVRKKQBQLYLWSTGGFTTALPDIV---NTDGRAATTPAETENAEW 183
Db      179 ITMLYS--GNAYDL SQVCCLAYAVNS-----DFLLIEWKKYRGNPVLLPPPVGVGKD 229
QY      184 FRDPKHWDTARGEWCVCIG----RLRYAAFYTSPNLRDWTLRRNFDPNHALLGGECP 238
Db      230 FRDPSTLWSGPDGEYRWWMGSKHNKTICCALLIYHTINTHFELKEEVLUHAVPHGTGMWECV 289
QY      239 DLFEIT-----ADGTRH-WVLAASMDA-----ADGTRH-WVLAASMDA-----YG 262

```



Qy 305 S---IDAPETKRLATAW-----MNNWKY-AARDVPTDASGYNGQNSIIVRELRLAROP 353  
Db 401 RTVLVLDL-ETKTNLQWPLEETENLSKKYDEFKDV-----ELR-----P 439  
Qy 354 GGWYLLSTPVAALNY-VTATTLLPDRTVGSAVLPHNGRAYEIELDIANDTATNVGIS 412  
Db 440 G---ALVPLEIGTATQOLDIVATFEIDQKMLEST-----LEADVLENCFTTSEG-S 484  
Qy 413 VGRSPDGRHRTNIGKYGADLYVDRGPS-----DLAG-----YSLAPYSR----- 451  
Db 485 VARSV-----LGFGVVVLADAGORSEQLPVYFIAKDIDGTSRTYFCADETRSSKDV 537  
Qy 452 -----AAAPIDPGARSVHLRLVDTQSVFVFNAGHTVLSQQVH 490  
Db 538 VGKWWYGVSSVPLPCEK-YNMRLLDVHSIVEGFAQGRVTVTSRYV 582

## RESULT 12

US-08-860-091A-4

; Sequence 4, Application US/08860091A

; Patent No. 6057494

; GENERAL INFORMATION:

; APPLICANT: Koops, Dr. Andries Jurriaan

; APPLICANT: Van den Meer, Dr. Ingrid Maria

; APPLICANT: Van Tunen, Dr. Arjen Johannes

; TITLE OF INVENTION: DNA sequences encoding carbohydrate

; TITLE OF INVENTION: polymer synthesizing enzymes and method for producing transgen

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LADAS &amp; PARRY

; STREET: 26 WEST 61 STREET

; CITY: NEW YORK

; STATE: NY

; ZIP: 10023

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/4" Floppy disk 1.44MB

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Microsoft Windows for Workgroups

; SOFTWARE: WordPerfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/860,091A

; FILING DATE: 02-JULY-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/NL/00012

; FILING DATE: 08-JANUARY-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Cord, Janet I.

; REGISTRATION NUMBER: 33,778

; REFERENCE/DOCKET NUMBER: U-011329-3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 708-1935

; TELEFAX: (212) 246-8959

; TELEX: 233288

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 615 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-860-091A-4

Query Match 6.7%; Score 189; DB 3; Length 615;

Best Local Similarity 20.0%; Pred. No. 9e-09;

Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHGHAYQLYLLHSDQNN--GPGGWDHASTTDCGVAFTTHGT 97  
Db 88 RTAFHFQPAKFIYDPDGLFHMGWYHMFYQNPYPVPMNMSWGHVSFKDMINNYE--- 144  
Qy 98 VMLPRDPFVMSGSAVGTANTAGGAGAVVALATQPTDGVKRYEQVLYLWSTDCGFTFT 157  
Db 145 -LPVAVPTEWYDIEGVLSGSTTVPNGQIFALYTGANDFSQLQ-----CK 190

Qy 158 ALP-----DPVIVN-----TDGRAATTPAETENAEFRDKIHWDTARGEWVCVIGRLR--- 206  
Db 191 AVPVNLSDPLLIEWKYEDNPILYTPPGI-GLKDYRDPSTVWTGPDGKHRMIMGTKRGNT 249  
Qy 207 --YAAFYTSPNLRDWTLLRNFDYPNHALGGI---BCPDLFETADDTGRHWHVLAASMDAY 261  
Db 250 GMVLVYTT---DYNTYELDELPHSVPTNDMWCVDYFPVSLTND---AIDMAAY 300  
Qy 262 GIGLPMYAYWTGWDGEQFHADDLTQWLMDGWMDWYAATWPSI-----DAPE----- 310  
Db 301 GSGIKHVI---KESWEGH-----GMDWYSIGTYDAINDKMTDPNDELVDGI 343  
Qy 311 -----TKRLAIAMNNWKYAADVPTDASDGYNGQNSIIVRELRLARQ 352  
Db 344 GLRCDYGRFFASKSLYDPLKKRRITW-----GYVGESDSA-DODLSR- 384  
Qy 353 PGGWYLLSTPVAALTNVYVTTTLPDRTVGSAVLPM-----NGRAYE----- 396  
Db 385 --GWATV-----YNVGRTIVLDRKT-GTHLLHWPVEEVESLRYNGQEFKEIKLEP 431  
Qy 397 ---IELDIAWDATNVGI-----SVGRSPDGTGRHTNIGKY 428  
Db 432 GSIIPLDIG--TATQLDIVATPEVDQAAALNATSETDDIYGCTTSLGAAQRG---SLGPF 485  
Qy 429 GADLYVDRGPSDLAGYSLAPYSRAAAPIDPGARSVH-----N----- 464  
Db 486 GLAVLADGTLSLT---PVFYIAKKADGGV-STHFCTDKLRSSLDYDGERVVYGGTVP 540  
Qy 465 -----LRILVDTQSVFVFNAGHTVLSQQVH 490  
Db 541 VLDDEELTMRLLVDHSIVEGFAQGRVTITSRAY 574

## RESULT 13

US-09-245-323A-8

; Sequence 8, Application US/09245323A

; Patent No. 6365800

; GENERAL INFORMATION:

; APPLICANT: Caimi, Perry G.

; TITLE OF INVENTION: Transgenic Crops Accumulating Fructose Polymers and  
; TITLE OF INVENTION: Methods for Their Production

; FILE REFERENCE: BB1082 US NA

; CURRENT APPLICATION NUMBER: US/09/245,323A

; CURRENT FILING DATE: 1999-02-05

; PRIOR APPLICATION NUMBER: 60/077,727

; PRIOR FILING DATE: 1998-03-12

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: MICROSOFT OFFICE 97

; SEQ ID NO 8

; LENGTH: 615

; TYPE: PRT

; ORGANISM: Helianthus tuberosus

US-09-245-323A-8

Query Match 6.7%; Score 189; DB 4; Length 615;

Best Local Similarity 20.0%; Pred. No. 9e-09;

Matches 115; Conservative 70; Mismatches 179; Indels 210; Gaps 27;

Qy 40 RAVYHMTPPSGWLCDPQRPVTHGHAYQLYLLHSDQNN--GPGGWDHASTTDCGVAFTTHGT 97  
Db 88 RTAFHFQPAKFIYDPDGLFHMGWYHMFYQNPYPVPMNMSWGHVSFKDMINNYE--- 144  
Qy 98 VMLPRDPFVMSGSAVGTANTAGGAGAVVALATQPTDGVKRYEQVLYLWSTDCGFTFT 157  
Db 145 -LPVAVPTEWYDIEGVLSGSTTVPNGQIFALYTGANDFSQLQ-----CK 190

Qy 158 ALP-----DPVIVN-----TDGRAATTPAETENAEFRDKIHWDTARGEWVCVIGRLR--- 206  
Db 191 AVPVNLSDPLLIEWKYEDNPILYTPPGI-GLKDYRDPSTVWTGPDGKHRMIMGTKRGNT 249  
Qy 207 --YAAFYTSPNLRDWTLLRNFDYPNHALGGI---BCPDLFETADDTGRHWHVLAASMDAY 261  
Db 250 GMVLVYTT---DYNTYELDELPHSVPTNDMWCVDYFPVSLTND---AIDMAAY 300

Db 250 GMLVYVYTT-----DYTNVELLDPLHSVPNTDMWECVDFPVSLTND-----ALDMAAY 300  
QY 262 GIGLPMYAYTGTWDEGFHADDLPOMLDGWDWYAAVTPSI-----DAPE----- 310  
Db 301 GSGIKHVI-----KESWEGH-----GMDWYSIGTYDAINDKWTDPNPELDVGI 343  
QY 311 -----TKRLAIAMNNKYYAARDVPTDASGYNQNSIVRELRLARQ 352  
Db 344 GLRCDYGRFFASKSLYDPLKKRITW-----GYVGESDSA-DQDLRS- 384  
QY 353 PGGWYLLSTPVAALTYVATTTLDPRTVDCSAVLPW-----NGRAYE----- 396  
Db 385 ---GWAIV-----YNGRTIVLDKRT-GTHLLHWPVEVESURYNGOEFKEIKLEP 431  
QY 397 ---IELDIAMDTATNGVI-----SVGRSPDGTGTRHTNIGRY 428  
Db 432 GSIPLDIG--TATOLDIVATFEVQOALNATSETDDIYGCITSLGAAORG-----SLGPF 485  
QY 429 GADLYVDRGPSDLAGYSLAPYSRAAIPDPGARSVH----- 464  
Db 486 GLAVLADGSLSELT-----PVIFYIAKKADGGV-STHFCTDKLRSLDYDGERVYVGGTVP 540  
QY 465 -----LRLVDTQSVFVFNAGHTVLSQQVH 490  
Db 541 VLDDELTMRLLVDHSIVEGFAQGGRTVITSRAY 574

RESULT 14  
US-09-019-385-2  
Sequence 2, Application US/09019385  
Patent No. 6147280  
GENERAL INFORMATION:  
APPLICANT: Smeekens, Josephus C.M.  
APPLICANT: Ebskamp, Michael J.M.  
APPLICANT: Geerts, Hendrikus A.M.  
APPLICANT: Weisbeek, Petrus J.  
TITLE OF INVENTION: Production of Oligosaccharides in  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson & Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,385  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/479,470  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: NL 1000064  
FILING DATE: 05-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: ARNO18589  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-224-0709  
TELEFAX: 206-224-0779  
TELEX: 4938023  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 626 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-019-385-2  
Query Match 5.9%; Score 166.5; DB 4; Length 626;  
Best Local Similarity 21.7%; Pred. No. 1e-06;  
Matches 135; Conservative 76; Mismatches 233; Indels 177; Gaps 36;  
QY 16 AGALALIFGAV-----PPAARASAPGS-----LRAVYHMTPPSGWLCDPQR 57  
Db 40 ASAMAVVVVGATLLAGLRMEQAVDEAAAGGFPWSNEMLOWQSGYHFOTAKNYSDPNG 99  
QY 58 PVTTHGAYGLYLYHSDQNNPG-GMDHASTTDGVAFTH-----HGTVMPLRPDPFVW- 108  
Db 100 LMYIRGWYHMFY---QYNPVGTDMD-----DGMENGHAVSRNLVQWRTLPIAMVADQWY 150  
QY 109 -----SGSAVV---CTA---NTAGFGAGAV---VALATOPTDGV-----RKYOEQYLYW 148  
Db 151 DILGVLSSGSMTVLPNGTVIMITYTGATNASAVEVQCITATPADNDPLLRWTKHPANPVIW 210  
QY 149 STDGGFTTALPOPPIVINTDGRAATTAPAEIENAEWFRDKPIHW-DTARGEWVCVIGRL-- 205  
Db 211 SPPG-----VGTKD-----PRDPTAWYDESDETWTLLGSKDD 244  
QY 206 -----RYAAFTSPNLRDWTLRNFDYPNHALGGIECDLFEI---TADCGTR--HWVL 254  
Db 245 HDGHHGDIAMMYKTDFLNYELIPGILHRVVRTGEWECIDFYVGRSSDSSSEMLH-VL 303  
QY 255 AASM-----DAYGIGLPMYAYTGTWDEGFHADDLPQWLDWG---WDW---YAAVT 302  
Db 304 KASMDDERHDYSLG---TYDAAANTW-----TPIDPE-LDLGIGLRYDMGKFYASTS 352  
QY 303 WPSIDAPETKRLAIAMNNKYYAARDVPTDASDYGNGQNSIVRELRLARQPGWGTLLST 362  
Db 353 F--YDPAKNERVLMGYGVEVDSKRAVD---VKGWASIQSVPTVALDEKTR--INLLW 404  
QY 363 PVAALTNYVTATTLDPRTVD--GSAY-LPWNGRAY-----EIELDIAMDWTATN---VGI 411  
Db 405 PVETETLRNATLTDVTTINTGSIHPIRQGTTHARHAESFHLDDASAVAALNEADVGY 464  
QY 412 SVGRSPDGTGTRHTNIGKYGADLYVDRGPSDLAGYSLAPYSRAAIPDPG----- 459  
Db 465 NCSSSGGAVNRGALGPFG-LVLAAG--DRRGEQTAVFYVSRGLDGLGHTSFCODELRS 521  
QY 460 -----ARSHLRILVDTSQSVFVFNAGHTVLSQQVHFAEG--DTGI 498  
Db 522 SRAKDVTKRVIGSTVPLDGEALSMRVLVDHSIVQGFDMGGRTTTSRYYPHESYQEARV 581  
QY 499 SLYTDGGPAHFTG--IVVREI 517  
Db 582 YLFNNATGASVTAERLVVHEM 602

RESULT 15  
US-09-142-623-11  
Sequence 11, Application US/09142623  
Patent No. 6337201  
GENERAL INFORMATION:  
APPLICANT: Koji YANAI et al.  
TITLE OF INVENTION: -FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
TITLE OF INVENTION: ISOLATING -FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING  
TITLE OF INVENTION: -FRUCTOFURANOSIDASE, AND -FRUCTOFURANOSIDASE VARIANT  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/142.623  
FILING DATE: September 10, 1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 98-0989\*/LC(WMC)/144  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 565 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: No. 6337201 relevant  
TOPOLOGY: No. 6337201 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Penicillium roqueforti IAM7254  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..565  
IDENTIFICATION METHOD: E

Search completed: May 6, 2003, 12:36:35  
Job time : 22.4625 secs

Query Match 5.2%; Score 147.5; DB 4; Length 565;  
Best Local Similarity 21.8%; Pred. No. 4.8e-05;  
Matches 126; Conservative 68; Mismatches 226; Indels 159; Gaps 29;

QY 26 AVPPAARASAPCSL---RAVYHMTPPSCWLCDP---QRPVTHHCAYQLYVYVHSDQNG 77  
DB 11 SAPPMLSTLANASLPTKWRPRAHLPLPPSGNIGDPCGHYTDPKT--GLFHVGLYS-----G 64  
QY 78 PGGMHASTTDCGVAFTHGTVMLRPDPFVWGSAGVAVGTAN--TAGFGAGAVV--ALATQP 134  
DB 65 ISG---ATTDDL---TYKDLNPD---GAPSIAGKNDPLSVFDGSLVPSIDGMP 112  
QY 135 T-----DGVRYQEQYLYWSTDGGTFTALPDVIVNTDGRAATTPAEIE 179  
DB 113 TLLYTSVSYLPIHWSIPTYTRGSETSLAVSDGHNFTKL-----NOGPVPTPPEAL 165  
QY 180 NAEWFRDPKIH---DTARGEWCVIG-----RLRYAAFTYSPNLRD 218  
DB 166 NTAERDPYVFPSPILDKSVNSTQGTWTVVAISGGVHGVGCPQFLYRQNDADFYWEYLQ 225  
QY 219 W---TLRNFDPNHALG---GIECPDLFEITADDGTRHWLAAASMDAYGIGLP----- 266  
DB 226 WKKEPLNTTWKGDWAGGWFNFVGNVFLNAEGYSEDGEIFITLGAEGSLPIVQVS 285  
QY 267 ---MTYAYWTGTWGDGFHADDLPQWLDGMDWYAAV--TWPSIDAPETKR-----L 314  
DB 286 STRDLWLTGNTVNDGVSFTKPTMAGV-LDMGVSAAYAAAGKILPASQASTKSGAPDRFI 344  
QY 315 AIAWNNKYY-AARDVPTDASDYGNSIVREL-----RLARQ 352  
DB 345 SYVWLTGDLFEQVGKFT- AQONWTGALLPRELNVRTISNVVDNELSRESLTWSRVARE 403  
QY 353 PGGWYTLTSTPVAALTNTVTAFTT-----LPDRTVDGSAVLFPN---GRAYEIELDIAW 403  
DB 404 DSGQIDLETMGISISRETYTSALTSGSSVFESGKTLNAGAVFPNTSPSKFFVLTANISF 463  
QY 404 DT-ATNVGISVGRSPDGRTHNIGKY----GADLYVDRGPSDLAGYSLAPYSRAAPI-- 456  
DB 464 PTSARDSGTQAGFQVLSSLSLESTTIYQFSNESIIVDRSNTSAA-----ARTTAGILS 516  
QY 457 --DPG-----ARSHLRILVDTQSVFVFN 479  
DB 517 DNEAGRLRLFDVLRNGKEQVETLELTIVVDNSVLEVYAN 555



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:34:52 ; Search time 17.4663 Seconds  
(without alignments)  
2563.947 Million cell updates/sec

Title: US-09-868-328B-4  
Perfect score: 2837  
Sequence: 1 MTPAIRRRRAVLQAGAGALA.....LYTDGPAHFTGIVVREIGQ 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 499.5 | 17.6        | 943    | 10 | US-09-969-362-5    |
| 2          | 468.5 | 16.5        | 1277   | 10 | US-09-969-362-3    |
| 3          | 461   | 16.2        | 923    | 10 | US-09-969-362-1    |
| 4          | 221   | 7.8         | 635    | 9  | US-10-003-392-14   |
| 5          | 206   | 7.3         | 630    | 9  | US-10-003-392-16   |
| 6          | 205.5 | 7.2         | 630    | 9  | US-10-003-392-18   |
| 7          | 202.5 | 7.1         | 433    | 9  | US-09-738-626-6405 |
| 8          | 189   | 6.7         | 615    | 9  | US-10-003-392-17   |
| 9          | 184   | 6.5         | 608    | 9  | US-10-003-392-2    |
| 10         | 173   | 6.1         | 625    | 9  | US-10-003-392-21   |
| 11         | 157   | 5.5         | 600    | 9  | US-10-003-392-20   |
| 12         | 155   | 5.5         | 609    | 9  | US-10-003-392-4    |
| 13         | 147.5 | 5.2         | 565    | 9  | US-09-990-385-11   |
| 14         | 141.5 | 5.0         | 635    | 9  | US-09-990-385-1    |
| 15         | 140.5 | 5.0         | 635    | 9  | US-09-990-385-13   |
| 16         | 131.5 | 4.6         | 498    | 9  | US-09-738-626-4737 |
| 17         | 130.5 | 4.6         | 495    | 9  | US-10-003-392-8    |
| 18         | 126   | 4.4         | 390    | 9  | US-10-003-392-6    |
| 19         | 123   | 4.3         | 471    | 9  | US-10-003-392-10   |

Sequence 171, Appl  
Sequence 33, Appl  
Sequence 33, Appl  
Sequence 33, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 1, Appl  
Sequence 136, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 227, Appl  
Sequence 2, Appl  
Sequence 4, Appl  
Sequence 31, Appl  
Sequence 31, Appl  
Sequence 31, Appl  
Sequence 7, Appl  
Sequence 6, Appl  
Sequence 11, Appl  
Sequence 6888, Ap  
Sequence 14, Appl  
Sequence 15, Appl  
Sequence 16, Appl

US-09-969-362-5  
; Sequence 5, Application US/09969362  
; Patent No. US20020076790A1  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rahbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And  
; FILE OF INVENTION: Process For Using The Enzyme  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/969,362  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: 09/397,885  
; PRIOR FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: PA 1998 01623  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: 60/101,615  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/111,675  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-09-969-362-5

Query Match.  
Best Local Similarity 30.8%; Pred. No. 6.8e-33;  
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;

Oy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNNNGPGWMDHASTTDDGVAFTHHGTVM 99  
Db 383 RPDYHSPARCSADPNGLVYEGYHLF--HQD-----GCTWAHAVSTDLVHWKRLPTAL 436  
Oy 100 PLRPDPFWMSGSAVGTANTAGF-----GAGAVVALATQPTDGVKRYQEQVLYWSTDDGGF 154  
Db 437 PWNDLGHWSGSAAVDLHNASGLFADSGKGLIAYTTSYNPDPRNGNRIGLAYSKDGR 496

#### ALIGNMENTS

RESULT 1  
US-09-969-362-5  
; Sequence 5, Application US/09969362  
; Patent No. US20020076790A1  
; GENERAL INFORMATION:  
; APPLICANT: Moller, Soren  
; APPLICANT: Johansen, Charlotte  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Ostergaard, Peter Rahbek  
; APPLICANT: Hoeck, Lisbeth Hedegaard  
; TITLE OF INVENTION: A 2,6,-b-D-Fructan Hydrolase Enzyme And  
; FILE OF INVENTION: Process For Using The Enzyme  
; FILE REFERENCE: 5540.200-US  
; CURRENT APPLICATION NUMBER: US/09/969,362  
; PRIOR FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: 09/397,885  
; PRIOR FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: PA 1998 01623  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: 60/101,615  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/111,675  
; PRIOR FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Paenibacillus macerans  
US-09-969-362-5

Query Match.  
Best Local Similarity 30.8%; Pred. No. 6.8e-33;  
Matches 157; Conservative 67; Mismatches 223; Indels 63; Gaps 19;

Oy 40 RAVYHMTPPSGWLCDPQRPVTHGAYQLYLHSDQNNNGPGWMDHASTTDDGVAFTHHGTVM 99  
Db 383 RPDYHSPARCSADPNGLVYEGYHLF--HQD-----GCTWAHAVSTDLVHWKRLPTAL 436  
Oy 100 PLRPDPFWMSGSAVGTANTAGF-----GAGAVVALATQPTDGVKRYQEQVLYWSTDDGGF 154  
Db 437 PWNDLGHWSGSAAVDLHNASGLFADSGKGLIAYTTSYNPDPRNGNRIGLAYSKDGR 496

|     |    |     |  |                                  |     |
|-----|----|-----|--|----------------------------------|-----|
| 83  | Qy | 137 | HASITDGVAFTHHGCTVMPLEPDEPVRSGSAVWGTA     | TAGF-----GAGAVVALATQPTDC         | 137 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 434 | -  | 493 | HAVSDJLNKRLPIALPWNDDQGVHSGSAIDLN         | NASGLFTDSGGKGLIAYTYSYHPDK        | 493 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 138 | Qy | 195 | VRK7QEQYLYWSTDGGTFTF-ALPDPVIVNTD         | GRAATTPAEINAEW-FRDPRKIHWDTAR     | 195 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 494 | Db | 548 | PGGNORIGLAYSTDGGRNQYAKERPVIDNPK          | NGDDP-----GSWDFDPKVVREDDH        | 548 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 196 | Qy | 253 | GEWCVIGRLRYAAYFYSNPLRDMWTLRRNFDY         | PNHALGGI-ECPDLPFEITAD-DGTRHWV    | 253 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 549 | Db | 608 | NRWVMVYSGGDHIREFTSTNLLDWTLTDFN           | GYG DYVRGVWECPDILQLPVDGTGORKWV   | 608 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 254 | Qy | 305 | LAA5MDAYGICLPMTYA- ---YNTG- -TW          | DEGFPHADLLTPOML- -DWGWDWAAAVTWPS | 305 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 609 | Db | 663 | LLISTGAN- ---PKTQSGDAEYFVGQUTADG         | -RFLNDHPAGQVLRDITYGKEFTASMSFAN   | 663 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 306 | Qy | 365 | IDAPETRKLAIAMNNKNKYAARDVPTDASDGY         | NGQNSIVRELRLARQPGGWYTLTSTPVA     | 365 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 664 | Db | 716 | M- -PNQRKVMLAWNTNDYPF-EFPTSS- -          | WKQLTI PREVSLRTTDEG-VRLVQTPTIT   | 716 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 366 | Qy | 419 | ALTNYVTATTTLPDRTVDGSAVLPWNG- -           | RAYEIELDI- - -AMDATNNGVSGVSRPDG  | 419 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 717 | Db | 775 | ELOKRLHNLXSAQQMTYGPKSKNPLEGLTAG          | EYIEAEVIEPANSVTEFGRFOL-ROREQ     | 775 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 420 | Qy | 476 | TRHNTICKYGAD- - -LVVDGRPSDLAGYSL         | APYSLAPYSRAAAPIDPCARSVHLRLIVDTS  | 476 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 776 | Db | 833 | QKTT- -VAYRVDTONMVFDRTTSGDVSFSD          | LFTKVIEASLKPNQKVKLRLFVDESSVEV    | 833 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 477 | Qy | 514 | FVNAGHTVLSQOVHFAEGDGTGISLYTDGGP          | HAFTGIVV                         | 514 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |
| 834 | Db | 871 | FGNDGKVVFSVDVFPDPAGRAMAFYISLGG           | BVKVSSMKV                        | 871 |
|     |    |     | : : : :    : : : :    : : : :    : : : : |                                  |     |

; Patent No. US20020076790A1

; GENERAL INFORMATION:

APPLICANT: JOHANSEN, Charlotte

APPLICANT: Hoeck, Lisbeth Hedegaard

; TITLE OF INVENTION: PROCESS FOR U

; CURRENT APPLICATION NUMBER: US/09/

;; PRIOR APPLICATION NUMBER: 09/391,800  
: PRIOR FILING DATE: 1998-08-17

PRIOR FILING DATE: 1998-12-08

; PRIOR FILING DATE: 1998-09-24

PRIOR FILING DATE: 1998-12-10

```

; SOFTWARE: FASTCSEQ FOR WINDOWS VERS.
: SEQ ID NO 3

```

; TYPE: PRT

US-09-969-362-3

Query Match: 16.5%; Score: 20.14

14/ JANUARY 1964 1607

The gel electrophoresis image displays multiple lanes containing horizontal bands representing DNA fragments. The bands vary in intensity and position across the different lanes, indicating differences in the genetic material being analyzed.

**Figure 1**

380 FODTYMTEESMVYKKNYBBOYHYVEDI BO

1

Db 446 PWNDRHGWSSGAVADMNTNASLFGDSGKGLIAYTTFSPNCPNGNORIGLAYSKDQGR 505  
Qy 155 TPTALPD--PVIYNTDGRAATTPAEIENAEW--PRDPKIHWDTARGWVVCVIGRLRAAFYT 212  
Db 506 TWEYSKERPIVLENGKSN-----EAGNWDPRDPKVIREDNENRNVVWVSGDHRIFT 560  
Qy 213 SPNLKRWTLRRNFDYPNHALGI--PCDLEFETADGT--RHWVLAASNDAYGIGLPMF-- 268  
Db 561 STNLDDWTLTWNWGYDVGWVWBCPDLPV--DGTSSQKKWMMISTGAN-----PKTG 615  
Qy 269 ---YAYWTC--TWGEQFHADDLTQWL--DWGWDWYAAVTPWSDAPETKRLATAWNN 321  
Db 616 GSDAEYFIGHLTADG--KFVNDNPAKVLRTDFGKFEYASMSFANM--PDHRTVMAMWNTN 672  
Qy 322 WKYAADVPTDASDYGNGNSIVRELRLARQPCGWYLLSTPVAALTNYVATTTLPORT 381  
Db 673 WDPYPA--FPT---SNWKGELTIPREVLSTVTTEDG--IRWVQSPIKEL-----ESLRKPLYS 722  
Qy 382 VDGSAVLPWNGR-----AYEIELDI---AMDATNVGIVSGRSPDGTNRHTNICKYGA 430  
Db 723 ASNKSVPSSGNLLKGIISGAYEIEAEIETPSTVTEFGFNIREG--ANOKTVVGYKAS 780  
Qy 431 D--LYVDRCPDLACYSLAPYSRAAPIDPGARSVHLRLILVDTQSVFVNAGHTVLSOO 488  
Db 781 DSRMEVDRTASGETDFSNLFSKKHEAPTOMENNRIMRLVDESSVEAFGNDGKVVFSDV 840  
Qy 489 VHFABEGDTGISLYTDGGPAHFTGIIVREI 517  
Db 841 ITPDPASRAMSFYKGGVGNVSVLKVHQL 869

## RESULT 4

US-10-003-392-14  
; Sequence 14, Application US/10003392  
; Patent No. US20020170086A1

; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Caimi, Perry G.  
; APPLICANT: Stoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BBI463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003.392  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 14

; LENGTH: 635

; TYPE: PRT

; ORGANISM: Parthenium argentatum Grey

US-10-003-392-14

Query Match 7.8%; Score 221; DB 9; Length 635;  
Best Local Similarity 21.5%; Pred. No. 4.4e-10;  
Matches 126; Conservative 82; Mismatches 207; Indels 170; Gaps 29;

Qy 12 QCAGAGALALIFGGAVPPAARASAPGSL-----RAYVHMTPPSGWCLCDPQRPVTH 62  
Db 69 QCA-----ATTFTTQLPKIDMKRVPGELDSGADVQWQRSAYHFQPKNVISDPDGPYHM 123  
Qy 63 GAYQLYLHSDQNN--GPGWDHASTTDGVAFTTHGTVMP--LRDPFVWSSAVVGTAN 118  
Db 124 GWYHLFYQYNPESAIWGNITWGHVSVDKMINWFH----LPFAVDPH--WYDIEGVMTGS 177  
Qy 119 TAGFCAGAVVALATOPTDGVKRYEQYLYWSTDGGFTETALPDPIV---NTDGRAATTP 175  
Db 178 ATVLPNGEIMLYT--GNAYDLSQVQCLAYAVNSS-----DPLLEWKKYEGNPVLLP 228  
Qy 176 AEIENAEWFRDPKIHWDTARGWVVCVIG-----RLRYAAFTYSPNLRDWTLRRNFDYPNH 230

Db 229 PPGVGYKDFRDPSTLWLGPDGEYRMVMSKHNHETGICALIYHTTNTHFELNEEVLHAYP 288  
Qy 231 ALGGTECPDLFEIT-----ADDGTR--HWVLAASMDA-----YGIGLPMTYAYWTGT 275  
Db 289 HTGMWECVDLYPVSTHTNGLDMVNDGPNVKYLVKQSGDEDRHDWYAIG-----SYD 340  
Qy 276 WDGEQFHADDLTPO-----WLDWGWDWYAAVTPWSDAPETKRLATAWNNWKAARD 328  
Db 341 WYNDKWPDD--PENDVGIGRLRYDG--KFYASKTF-----YDQHKRRVLW----- 383  
Qy 329 VPTDASDYGNGNSIVRELRLARQPCGWYLLSTPVAALTNYVATTTLTLP----- 378  
Db 384 -----GYVGEDT--PEKYDLTK--GWANILNIPRTVVLDTKTKNLNLIQWPIETEKLR 432  
Qy 379 ---DRTVD-----GSVLPWNGRAVE-----IELDITAMDTATNVGTS 412  
Db 433 SKYDKFVDEVELRPGSLPLEIGTATQLDIVATFEVDQMMLLESTLEADVLNCTTSVG--S 491  
Qy 413 VGRSPDGTNRHTNICKYGADLYVDRG-----PSDLAG-----YSLAPYSRAAAPID 457  
Db 492 VGRGV-----LGPFGVVVLADAQRTQOLPVYFYIAKDTDGTSTRTYFCADETRSSKVD 544  
Qy 458 PGA-----RSVHLRLILVDTQSVFVNAGHTVLSOOVH 490  
Db 545 VGVWYGVSSVPVLPNEKYNMRLLDVHSIVEGPAQNGRTVVTSRVY 589

## RESULT 5

US-10-003-392-16  
; Sequence 16, Application US/10003392  
; Patent No. US20020170086A1

; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Caimi, Perry G.  
; APPLICANT: Stoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BBI463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003.392  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16

; LENGTH: 630

; TYPE: PRT

; ORGANISM: Helianthus sp.

US-10-003-392-16

Query Match 7.3%; Score 206; DB 9; Length 630;  
Best Local Similarity 21.8%; Pred. No. 7.6e-09;  
Matches 125; Conservative 81; Mismatches 204; Indels 164; Gaps 29;

Qy 28 PPAARASA-----PGSL-----RAYVHMTPPSGWCLCDPQRPVTHGAYOL 67  
Db 62 PPAATFTATQLPNIDLKRVPGKLDSSAEVWQRSAYHFQPKNFISDDPGPMYHGWHL 121  
Qy 68 YLHSDQNN--GPGWDHASTTDGVAFTTHGTVMP--LRDPFVWSSAVVGTANTAGFG 123  
Db 122 FYQYNPESAIWGNITWGHVSVDKMINWFH----LPFAVDPH--WYDIEGVMTGSATVLP 175  
Qy 124 AGAVVALATOPTDGVKRYEQYLYWSTDGGFTETALPDPIV---NTDGRAATTPAEIEN 180  
Db 176 NGQIIMLYT--GNAYDLSQVQCLAYAVNSS-----DPLLEWKKYEGNPVLPFGPVG 226  
Qy 181 AEWFRDPKIHWDTARGWVVCVIG-----RLRYAAFTYSPNLRDWTLRRNFDYPNHALGI 235  
Db 227 YKDFRDPSTLWLGPDGEYRMVMSKHNHETGICALIYHTTNTHFELKEEVLHAYPHTGMW 286  
Qy 236 ECPDLFEIT-----ADDGTR--HWVLAASMDA-----YGIGLPMTYAYWTGTWDEQ 280

Db 287 ECVDLYPVSTVHTNGLMDVNGRNKVLKQSGDEDRHDWYAG-----SYDVVNDKWYVDD 343  
QY 281 FHADDLTPQWLMDGWYAAVTPSIDAPETKRLATAMNNWYAAARDVPTDASDYGNGQ 340  
Db 344 PENDVGIGLYDFG-KFYASKTF-----YDQHKRRVLW-----GYVGE 381  
QY 341 NSIVRELRLARQPGGWYLLSTPVAAL-----TNY-----VTATTTLPDRTVD----- 383  
Db 382 TD-PQKYDISK---GWANILNIPRTVVLDTKTNTLNQIOWPIETENLRSKTYDEFKDVEL 437  
QY 384 --GSVLPWNGRAYE-----IELDIAMDTATNVGISVGRSPDGRHTN 424  
Db 438 RPSGLVPLEIGTATQLDIVATFEIDOKMLESTLEADVLFNCITTSSEG-SVARGA----- 489  
QY 425 IGRYGADLYVDRGFS-----DLAG-----YSLAPYSR-----AAAPI 456  
Db 490 LGPEGVVVLADARSEQLPVYFIANDIDGTSRTYFCADETRCKOVSVGKWKYVSSVPV 549  
QY 457 DPGARSVHLRLVDVTSQVSVFVFNAGHTVLSQQVH 490  
Db 550 LPGEK-YNMRLLDVHSDIVEGFAQNGRTVVTSRVY 582

## RESULT 6

US-10-003-392-18  
; Sequence 18, Application US/10003392  
; Patent No. US20020170086A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Calimi, Perry G.  
; APPLICANT: Scoop, Johan M.  
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes  
; FILE REFERENCE: BB1463 US NA  
; CURRENT APPLICATION NUMBER: US/10/003,392  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/244,273  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: 60/269,543  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 18  
; LENGTH: 630  
; TYPE: PRT  
; ORGANISM: Helianthus tuberosus  
US-10-003-392-18

Query Match 7.2%; Score 205.5; DB 9; Length 630;

Best Local Similarity 21.2%; Pred. No. 8.3e-09;  
Matches 124; Conservative 77; Mismatches 202; Indels 183; Gaps 29;

QY 20 ALIFGGAVPPAARASAPGSL-----RAVYHMTPPSGWGLCDPQRPVTHGAYOLYYL 70  
Db 65 ATTFVTQLPNIDLRKVPGLDSSAEVWQSTYHFQDKNFISDPDGPYHMGWYHLFYQ 124  
QY 71 HSDONN--GPGGWDHASTTDGVAFTHGVTMP--LRPDPVWGSAGVGTANTAGFCAGA 126  
Db 125 YNPQSAIWNITWGHVSYSKDMINWFH-----LPEAMVPDH--WYDIEGVTGTSATVLPNGQ 178  
QY 127 VVALATQPTDGVKRYQEQYLYHNSDGGFTTALPDVIV--NTDGRAATTPAEIENAEW 183  
Db 179 IIMLYS--GNAYDLSQVQCLAYAVNSS-----DPLLEWKYEGNPNVLLPPPGVGYKD 229  
QY 184 FRDPKIHWDTARGWCVCVIG-----RLRYAAFTSPNLRDWTLLRRNFDPNHALGEGICP 238  
Db 230 FRDPSTLWSCPDGEYRWVWGSKINETIGCALIYHTTNTFTHFELKEEVLHAPVHTGWECV 289  
QY 239 DLFEIT-----ADDDGTRH--WVLAASMDA-----YCG 262  
Db 290 DLYPVSSTVHTNGLMDVNGPNVKYVLKQSGDEDRHDWYAGSYDIYNDKWYVDDPENDVG 349  
QY 263 IGLPNTYATWTGTDGEQFHAD-----DLTPQWLMDGWNDWYAAVTWP 304

Db 350 IGL-----RYDFGKFYASKTFYDQHKRRVLNMGYVGETPOKYDLSKGSWANILNIP 400  
QY 305 S-----IDAPETKRLAIAM-----MNNWKY--AARDVPTDASDYGNGSIVRELRLARQP 353  
Db 401 RTVVLDL-EKTNLIQWPIETENLRSKKYDEPKDV-----ELR-----P 439  
QY 354 GGYTLLSTPVAALTNV--VTATTTLPDRTVDGSAVLPMNGRAYEIELDIAMDTATNVGIS 412  
Db 440 G-----ALVPLEIGTATQLDIVATFEIDOKMLEST-----LEADVLFNCITTSSEG-S 484  
QY 413 VGRSPDGRHTNICKGADLYVDRGFS-----DLAG-----YSLAPYSR----- 451  
Db 485 VARSV-----LGPEGVVVLADARSEQLPVYFIANDIDGTSRTYFCADETRCKOVSV 537  
QY 452 -----AAAPIDPGARSVHLRLVDVTSQVSVFVFNAGHTVLSQQVH 490  
Db 538 VGKWKYVSSVPVLPGEK-YNMRLLDVHSDIVEGFAQNGRTVVTSRVY 582

## RESULT 7

US-09-738-626-6405  
; Sequence 6405, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 6405  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6405

Query Match 7.1%; Score 202.5; DB 9; Length 433;

Best Local Similarity 24.8%; Pred. No. 9e-09;  
Matches 132; Conservative 49; Mismatches 166; Indels 185; Gaps 32;

QY 38 SLRAVYHMTPPSGWGLCDPQRPVTHGAYOLYLYHSDQNNNGPG-----GWDHAST--- 86  
Db 11 SLRPAHYVTPQGRNDPNMGVYDGTLLHYIQHD-----PGFPFAPKRI;GWAHTTPTLT 65  
QY 87 -TDGVAFTHGVTMP--LRPDP-----VWGSAGVVTANTAGGAGAVVALATQPTDGV 138  
Db 66 GPORLQWTH-----LPDALYPDASDYLDGCGYSGGAVFTDGTLLKLYTGNL-----KIDGX 115  
QY 139 RKQEOYLYWSTD-----GGFTTALPDVIVNTDGRAA--TTPAEIENAEWFRDPKIHWD 192  
Db 116 RR-ATONLVEVEDPTGLMGGIHRRSPKNPLI---DGPASGFTP-----HYRDPMLSPD 164  
QY 193 TARGE-WVCVIGRLR-----YAAFTSPNLRDWTLLRRNFY-----PNHALGGI- 235  
Db 165 ---GDGKWKVLGNQRENLTGAAYLYRSTDLNWEFSGEITFDLSDAQPSAPDLVPGGYM 221





|    |     |     |  |           |     |
|----|-----|-----|--|-----------|-----|
| Qy | 396 | EIE | -----LDIAWDTATNVGISGRSPDGTR-----HTN-----             | -----IGRY | 428 |
|    |     |     |  |           | :   |
| Db | 419 | EIE | PGSIIPLDIGSATQLDIVATFEVDQDALKAISEEYICTKSWGAAGRGSIGPF | 478       | :   |
| Qy | 429 | GAD | YVDRGRSPDLAGY-----SLAPYSR-----                       | AAAPIDPGA | 460 |
|    |     |     |  |           |     |
| Db | 479 | GVA | LDGTLSELTPVYFIAKNTDGSVATFCTDKLRSLDYDRRVYGVSTVPVLDG-  | 537       |     |
| Qy | 461 | RSV | RLVDVTQTSVEVEFVNAGHTVVLQQVH                          | 490       |     |
| Db | 538 | EEL | TRLLVDHSVVEGFAOGRVITSRVY                             | 567       |     |

```

RESULT 10
US-10-003-392-21
; Sequence 21, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calimi, Perry G.
; APPLICANT: Stoop, Johan N.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-003-392-21

```

```
Qy   413 VGRSPDGTRHNTICKYGADLYVDVRGSDLAGYSLAPYSRAAAIPDPG----- 459
      |         :|       |     ||    |:||        :||
Db   465 CSSGGAVNRGALGPFGI-LVLAAG--DRRGQTAVFYVSRLDGLHTSFCDLRRS 521
      |         :|       |     ||    |:||        :||
Qy   460 -----ARSHLRILIVDTQSVEFVNAGHTVLSSOQHFAEG--DTGIS 499
      :::::||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db   522 RAKDVTKRVIGSTVPVIDLGSAIMRVLVDHSIVOFGDMGGRTTMSKVPNIESYQEARY 581
      |         :|       |     ||    |:||        :||
Qy   500 LYTDGGPAHFTEG--INVREI 517
      |         :|       |     ||    |:||        :||
Db   582 LFNNATGASVTAEARLVVHEM 601
      |         :|       |     ||    |:||        :||

RESULT 11
US-10-003-392-20
; Sequence 20, Application US/10003392
; Patent No. US20020170086A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Stoop, Johan M.
; TITLE OF INVENTION: Fructan Biosynthetic Enzymes
; FILE REFERENCE: BB1463 US NA
; CURRENT APPLICATION NUMBER: US/10/003,392
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,273
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 60/269,543
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-003-392-20
```

Query Match



Db 166 NVTAFRDPYVQSPILDKSVNSTQGTWYVAISGVHGVGPCQFLYRONDADFQYWEYLQ 225  
QY 219 W---TLRRNFDYPNHALG---GTECPDLFEITADGTRHWVLAASMDAYGIGLP----- 266  
Db 226 WKEPLNTTWGKDWAGGNGFNEVGNVSLNAEGVSEDEIEITILGAEGLPIVPOVS 285  
QY 267 ---MTYAYWTGWGDEQFHADDLTQWLDWGDWYAAV--TWPSIDAPETKR-----L 314  
Db 286 SIRDMLVNTGNVNDGVSFTKPTMAGV-LDWGVSAYAAAGKILPASSQASTKSGAPDRFI 344  
QY 315 AIAWMNNKY-AARDVPTDASDGYNQNSIVREL-----RLARQ 352  
Db 345 SYVWLTGDFEQVKGFPT-AQQWNTGALLPRELNVRTISNVVDNELSRSTSWRARE 403  
QY 353 PGCWYLLTSPVAALNTYVATT---LPDRTVDCSAVLPWN---GRAYEIELDIAM 403  
Db 404 DSOQIDLEWNGISIRETSALTSFGSFSVSGKTLNAGAVPNTSPSSKFFVLTANISF 463  
QY 404 DT-ATNVGISVGRSPDGRHTNIGKY----GADLYYDRGPSDLAGYSLAPYSRAAAPI-- 456  
Db 464 PTSARDSGIQAGFOVLSLSLESTIYYQFSNESIIIVDRSNTSAA-----ARTTAGILS 516  
QY 457 --DPG-----ARSVHLRLVDTQSVFVN 479  
Db 517 DNEAGRLRLFDVLRNGKEQVETLELTIVVDNSVLEYIAN 555

RESULT 14

US-09-990-385-13  
; Sequence 13, Application US/09990385  
; Publication No. US20020192771A1  
; GENERAL INFORMATION:  
; APPLICANT: KOJI YANAI et al.  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCE  
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARI

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 Inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
FILING DATE: 10-Sep-1998  
PRIORITY APPLICATION DATA:  
FILING DATE: September 10, 1998

ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 2001-1611  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acid residues  
STRANDEDNESS: No. US20020192771A1 relevant  
TOPOLOGY: No. US20020192771A1 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Aspergillus niger ACE-2-1  
(ATCC 20611)

FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..635  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-990-385-1  
  
Query Match 5.0%; Score 141.5; DB 9; Length 635;  
Best Local Similarity 19.1%; Pred. No. 0.0017;  
Matches 125; Conservative 83; Mismatches 245; Indels 203; Gaps 28;  
  
QY 26 AVPPAARASAPGSL-----RAYHMTPPSGWLCDDPQRPVT--THGAYQLYLHSDQNGP 78  
Db 8 APPPTLSTLPNNTLPHVWRPRAHILPAEQIGDPCAHTDPSTGLFHVGLFL--DGDGI 65  
QY 79 GGDHAS-----TTDGVAFTHHGTVMPLRPDP--VWGSVAV-VGTANTAGGAGAVVA 129  
Db 66 AGATTANLATYDTSNGSF-----LIQGGKNDPVAVFDGAVIPGVNNTPTLLYTSVSF 121  
QY 130 LATQ-PTDGVKRYQEQLYWSTDGGFTF-----TALPDP----- 162  
Db 122 LPHWSIPVTRGETQSLAVARDGRRFDKLDQGPVIADHPFAVDVTAFRDPFVFRSAKL 181  
QY 163 -VIVNTDGRAATTPAEIENA-EWFRDPKIHWDTARGEWVCVIGRLRYAAFYTSPNLRD-- 218  
Db 182 DVLLSLDEEVARNETAVOAQAVDGVGTEKNAPWYVAVSGGVHGVGPAQFLYRQNGGNASEFQ 241  
QY 219 -----WTLRRNFDYPNHALGGI-----ECPDFLFEITADG 249  
Db 242 YWEYLGEWQEAATNSSWGDGTWAGRWGP---NFETGNVLFLEEGHDPQTGEVFTLGT 298  
QY 250 RHWVLAASMDAYGIGLPMT-----YAYWTTWGDGEQFHADDLTPQ---WLDWGW 295  
Db 299 E-----GSLUPIVPOVSSIHDMLEWAAAGEVGVGSEQEGAKVEFSPSMAGFLDWGF 347  
QY 296 DWYAAV--TWPSIDAPE-----TKRLAIAMNNKYYAAROVPTDASGYNQNSIVRE 346  
Db 348 SAYAAAGKVLPAASSAVSKTSKGVEDRVYFVWLTGQYEQADGFPPTAQOQGWTSLLPRE 407  
QY 347 LR-----LARQPG-----GWYTLSTPVAALNTYVATTTL 377  
Db 408 LKVQTVENVVDNELVREEGVSNVVGESDNOTARLRLTGITIARETKAALLAN--GSVTAE 465  
QY 378 PDRYDGSVAVLPW---NGRAYEIELDIAMDATN-----VGISVGRSFDGTRHT 423  
Db 466 EDTLTQAAVVPFAQSPSSKFFVLTAAQLBEPASARSPLQSGFEILASELERTAIYQFS 525  
QY 424 NICKYGADLYVDGRPSDLAGYSLAPYSRAAAPIDPG-----AR 461  
Db 526 N-----ESLVDRSQT-----AAAPTNPGLDSFTESGKLRLLFDVIENGEOQVE 569  
QY 462 SVHLRLVDTQSVVEFVNAGHTVLSQQVHFAEGDTGISLYTDG-GPAHFTGIIVRE 516  
Db 570 TDLTVVDNNAVVEYVANGRFALSTWARSWYDNTQIRFEHNGEVEGEVQFRNYSVSE 625

RESULT 15

US-09-990-385-13  
; Sequence 13, Application US/09990385  
; Publication No. US20020192771A1  
; GENERAL INFORMATION:  
; APPLICANT: KOJI YANAI et al.  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PROD  
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VA  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/990.385  
FILING DATE: 10-Sep-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/142.623  
FILING DATE: September 10, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: 2001-1611  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 574 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: No. US20020192771A1 relevant  
TOPOLOGY: No. US20020192771A1 relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Scopolariopsis brevicaulis IF04843  
FEATURE:  
NAME/KEY: mat peptide  
LOCATION: 1..574  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-990-385-13

Query Match 5.0% Score 140.5; DB:9; Length 574;  
Best Local Similarity 21.9%; Pred. No. 0.0018;  
Matches 129; Conservative 56; Mismatches 245; Indels 159; Gaps 29;

Qy 26 AVPPAARASAPGSL-----RAYVHMTTPSGWLCDDPQPVPT--THGAYQLYLLHSDQNNNGPG 79  
Db 19 SAPPNLSTLANNLSFETWRPRAHVLPPQNIQIGDCPMHYTDPETGIFHVGWLYN--GNGAS 76

Qy 80 GWDHASTDGVAFTHHGTVMPLRDPFVWGSAAV---GTANTAGFGAGAVV--ALATQP 134  
Db 77 G---ATTEDLV-----TYQDLNPD-----CAQMILPGGVNDPIAVFDGAVIPSGIDGKP 122

Qy 135 TDGVRYKQEOVLYW-----STDGGFTTAL-PDPVIVNTDGRAATTP--- 175  
Db 123 TMMTTSVSYMPISWSIAYTRGSETHSLAVSDGKFNFTKLQGVPIPPFPFGANVTSWRD 182

Qy 176 -----AEIENAEWF-----RDPKIHMDTARGEWCVIGR 204  
Db 183 PFLFQNFQFDSLLESENGTWTYVLSGGIHGDPSPAFLYRQHDPDFQYWEYLGPNWNEGN 242

Qy 205 LRYAAFTSPNLRDWTLRNFDYNHALGGIE---CPDLFEITADDGTRHVLAAASMDA 260  
Db 243 STWGS-----GDWAGRWGYNFEVINVLGDDGYNPD-GEIATVGT-EWSEFDPIKQP 293

Qy 261 YGIGLPWTYAWTCTWDEQFHADDLTPOWLGDWMDYAAV--TWPSIDAPETKR----- 313  
Db 294 ASDNREMLWAAGNMTLEDGDIKFTPSMAGYLDWGLSAYAAAGKELPASSKPSOKSGAPDR 353

Qy 314 -LATAWMNNKYYAARDPTDASDCYNGNSIVREL-----RLAROPGGWYTLIS 361  
Db 354 FVSYLNLITGDYFEGHDFET-PQQWNTGSLLLPRELSVGTIPNVVDNELARETGSW--RVG 410

Qy 362 TPVAALTYNYVT-----ATTTLPDRTVDS-----AVLPWNGRAYEIELDIATNATNVG 410  
Db 411 TNDTGVLVLTLYKQEIARETLAEMTSGNSTEASRNVSPPGTAFOQSDSKFFVLT-AS 469

Qy 411 ISVGRSPDGRHTWIGYAGDLYVDKPSDLGYSLA-----PYSRAAAPIDPG--- 459  
Db 470 LSF---PSSARDSL-RAGFEILSSEFSTVYQFSNESIIIDRSNSSAALATDGDIT 525

Qy 460 -----ARSVHLRLVDTQSVSEFVFNAGHTVLSQQV 489  
Db 526 RNEFGKMRFLFDVWEGDQERLETDLTIWVDSNIVEVHAN-GREALSTWV 573

Search completed: May 6, 2003, 12:44:13  
Job time : 20.4663 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 12:29:32 ; Search time 57.1096 Seconds  
(without alignments)  
1879.728 Million cell updates/sec

Title: US-09-868-328B-1

Perfect score: 2845  
Sequence: 1 MTPAISRAVLQAGAGALA.....TDGPAHFTGIVREIGQAI 521

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query % | Length | ID        | Description        |
|------------|--------|---------|--------|-----------|--------------------|
| 1          | 2845   | 100.0   | 521    | 2 Q9KJD0  | Q9KJD0 arthrobacte |
| 2          | 2355   | 82.8    | 517    | 2 O50585  | O50585 arthrobacte |
| 3          | 2335   | 82.1    | 530    | 2 Q9EV09  | Q9EV09 microbacter |
| 4          | 1047.5 | 36.8    | 620    | 2 Q93R69  | Q93R69 microbacter |
| 5          | 944    | 33.2    | 516    | 16 O07003 | O07003 bacillus su |
| 6          | 880.5  | 30.9    | 514    | 16 Q97I80 | Q97I80 clostridium |
| 7          | 659.5  | 23.2    | 395    | 2 P94469  | P94469 bacillus st |
| 8          | 562.5  | 19.8    | 512    | 2 O45372  | O45372 paenibacill |
| 9          | 527.5  | 18.5    | 501    | 2 Q9LAL1  | Q9LAL1 pseudomonas |
| 10         | 515    | 18.1    | 622    | 2 Q45155  | Q45155 bacteroides |
| 11         | 513    | 18.0    | 534    | 2 Q9RBJ1  | Q9RBJ1 acetobacter |
| 12         | 486.5  | 17.1    | 485    | 2 Q8RLU2  | Q8RLU2 xanthomonas |
| 13         | 469.5  | 16.5    | 493    | 2 Q9APV5  | Q9APV5 bacillus su |
| 14         | 468.5  | 16.5    | 493    | 2 Q93T55  | Q93T55 bacillus sp |
| 15         | 467.5  | 16.4    | 943    | 2 Q44109  | Q44109 actinomyces |
| 16         | 457    | 16.1    | 537    | 3 O42801  | O42801 aspergillus |

|    |       |      |      |           |                    |
|----|-------|------|------|-----------|--------------------|
| 17 | 448   | 15.7 | 537  | 3 Q96TU3  | Q96TU3 aspergillus |
| 18 | 433   | 15.2 | 516  | 3 O74641  | O74641 aspergillus |
| 19 | 427   | 15.0 | 516  | 3 O74642  | O74642 aspergillus |
| 20 | 427   | 15.0 | 516  | 3 O94220  | O94220 aspergillus |
| 21 | 425   | 14.9 | 515  | 3 O9HFA5  | O9HFA5 penicillium |
| 22 | 423   | 14.9 | 516  | 3 Q96W28  | Q96W28 aspergillus |
| 23 | 422.5 | 14.9 | 550  | 5 O02490  | O02490 tritricomo  |
| 24 | 421   | 14.8 | 516  | 3 O8X217  | O8X217 aspergillus |
| 25 | 416.5 | 14.6 | 533  | 3 O94224  | O94224 pichia jadi |
| 26 | 406.5 | 14.3 | 1142 | 16 Q97J24 | Q97J24 clostridium |
| 27 | 389.5 | 13.7 | 515  | 3 O00056  | O00056 penicillium |
| 28 | 386   | 13.6 | 750  | 2 O31411  | O31411 bacillus sp |
| 29 | 366.5 | 12.9 | 555  | 3 Q9UWF4  | Q9UWF4 kluyveromyc |
| 30 | 356   | 12.5 | 477  | 2 O86076  | O86076 escherichia |
| 31 | 355   | 12.5 | 477  | 16 O8XCLO | O8XCLO escherichia |
| 32 | 354.5 | 12.5 | 609  | 3 Q9V746  | Q9V746 kluyveromyc |
| 33 | 320.5 | 11.3 | 591  | 10 Q9ZTL2 | Q9ZTL2 zea mays (m |
| 34 | 313.5 | 11.0 | 453  | 3 O42878  | O42878 schizosacch |
| 35 | 304   | 10.7 | 513  | 5 Q9U0Y9  | Q9U0Y9 leishmania  |
| 36 | 304   | 10.7 | 670  | 10 Q8RVK8 | Q8RVK8 lolium pere |
| 37 | 302   | 10.6 | 812  | 2 Q9RBA5  | Q9RBA5 arthrobacte |
| 38 | 301   | 10.6 | 581  | 5 Q9U0Y8  | Q9U0Y8 leishmania  |
| 39 | 297.5 | 10.5 | 581  | 10 Q43856 | Q43856 vicia faba  |
| 40 | 295   | 10.4 | 695  | 3 Q9C214  | Q9C214 neopospora  |
| 41 | 285   | 10.0 | 661  | 10 Q94C05 | Q94C05 ipomoea bat |
| 42 | 283.5 | 10.0 | 580  | 10 Q43799 | Q43799 nicotiana t |
| 43 | 283   | 9.9  | 650  | 10 Q42722 | Q42722 daucus caro |
| 44 | 279   | 9.8  | 636  | 10 Q94114 | Q94114 prunus cera |
| 45 | 279   | 9.8  | 646  | 10 Q8RVH4 | Q8RVH4 cichorium l |

#### ALIGNMENTS

##### RESULT 1

| Q9KJD0 | ID   | Q9KJD0 | PRELIMINARY | PRT | 521 AA. |
|--------|--|--------|-------------|-----|---------|
| AC     | Q9KJD0   |        |             |     |         |
| DT     | 01-OCT-2000 (Tremblrel. 15, Created)                             |        |             |     |         |
| DT     | 01-OCT-2000 (Tremblrel. 15, Last sequence update)                |        |             |     |         |
| DT     | 01-MAR-2002 (Tremblrel. 20, Last annotation update)              |        |             |     |         |
| DE     | Levan fructotransferase.   |        |             |     |         |
| GN     | LFTA.  |        |             |     |         |
| OS     | Arthrobacter ureafaciens.  |        |             |     |         |
| OC     | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;          |        |             |     |         |
| OC     | Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.    |        |             |     |         |
| OX     | NCBI_TaxID=37931;  |        |             |     |         |
| RN     | [1]  |        |             |     |         |
| RP     | SEQUENCE FROM N.A.   |        |             |     |         |
| RC     | STRAIN-K-2032;   |        |             |     |         |
| RA     | Song K.B., Rhee S.K., Yoo E.J.;                                  |        |             |     |         |
| RT     | "Nucleotide sequence of levan fructotransferase gene (lftA) from |        |             |     |         |
| RT     | Arthrobacter ureafaciens K-2032."                                |        |             |     |         |
| RL     | Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.          |        |             |     |         |
| CC     | !- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.      |        |             |     |         |
| DR     | EMBL; AF181254; AAF73829.1;                                      |        |             |     |         |
| DR     | InterPro; IPR001362; GH_32.                                      |        |             |     |         |
| DR     | Pfam; PF00251; Glyco_hydro_32; 1.                                |        |             |     |         |
| KW     | Glycosidase; Hydrolase; Transferase.                             |        |             |     |         |
| SQ     | SEQUENCE 521 AA; 56590 MW; 863F3FFC4622D2B3 CRC64;               |        |             |     |         |

Query Match Best Local Similarity 100.0%; Pred. No. 1.7e-179; Length 521; Mismatches 0; Indels 0; Gaps 0;

|    |    |   |
|----|----|---|
| QY | 1  | MTPAISRAVLQAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPTPSGWLCDQRPVPT 60  |
| DB | 1  | MTPAISRAVLQAGAGALALIFGGAVPPAARASAPGSLRAYVHMTPTPSGWLCDQRPVPT 60  |
| QY | 61 | THGAYOLYLLHSDQNGPGWDHASTTDCGVAFTHTCTVMTPLRDPFVWSSGSAVGTANTA 120 |
| DB | 61 | THGAYOLYLLHSDQNGPGWDHASTTDCGVAFTHTCTVMTPLRDPFVWSSGSAVGTANTA 120 |

|  |   |                           |   |   |                 |         |
|--|---|---------------------------|---|---|-----------------|---------|
| Db   | 177   | AEWPRDPKIIHWDATRN         | EWCVIGIARYAAPTYS  | PNRLDQWQMSNFDP                          | PNHALGGTECPDL   | 233     |
| Qy   | 241   | FEITADDDGTRHVVLA          | ASMDAYGIGLPMYAY   | TWTGTDGQFADHDL                          | TLPQWLQWGDWDYAA | 300     |
| Db   | 237   | FEMTAGDGRHVVFGA           | SKDAYSIGLPMYFAY   | TGSGNGTAFIADNL                          | TLPQWLQWGDWDYAA | 296     |
| Qy   | 301   | VTWPSIDAPETKRLA           | TIAMNNWYKAARDV  | PTDASDGYNGONS                           | TVRELRLARPQGGWY | 360     |
| Db   | 297   | VTWPAVEAPEPKRLA           | TIAMNNWYKAARNV  | PTDASDGYNGQNS                           | TVRELRLERQSGGWY | 356     |
| Qy   | 361   | STPVAALNTYVATT            | PLPDRFVDGSAVLP  | WNGRAYEIELDIAW                          | DTATNNGVISVGRSP | 420     |
| Db   | 357   | STPVPALNSVAT              | STTLPDRFVNGS  | FVLPWMSGRAYEIE                          | LDSMTAANVGVSVGR | 416     |
| Qy   | 421   | RHTNIGKYGADLYD            | RGPSDLAGYSLAP   | YSRAAAPIDPGARS                          | VHLRLIVDTQSVEFV | 480     |
| Db   | 417   | RHTNIGKYGDELYD            | RASSQSGYALAPY   | TAAAPIDANARS                            | VHLRLIFVDTQSVEF | 476     |
| Qy   | 481   | GHTVLSQOVHPAGD            | TGIGISLYTGGPA   | HFTGIVVREIGAI                           | 521             |         |
| Db   | 477   | GHTVVSQOVHFAA             | GDGTGISLYADG  | GPANFTGITIREF                           | GNPI 517        |         |
| RESULT 3   |   |                           |   |   |                 |         |
| Qy   | Q9EV09  | ID                        | Q9EV09  | PRELIMINARY;                            | PRT;            | 530 AA. |
| AC   | Q9EV09  | DT                        | 01-MAR-2001   | (T+EMBLrel. 16, Created)                |                 |         |
| DT   | 01-MAR-2001   | DT                        | 01-MAR-2001   | (T+EMBLrel. 16, Last sequence update)   |                 |         |
| DT   | 01-MAR-2002   | DT                        | 01-MAR-2002   | (T+EMBLrel. 20, Last annotation update) |                 |         |
| DE   | Levan fructotransferase.  | GN                        | LFTM.   |   |                 |         |
| OS   | Microbacterium sp. AL-210.  | OC                        | Bacteria; Firmicutes; Actinobacteria;                   | Actinobacteridae;                       |                 |         |
| OC   | Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.  | ON                        | NCBI_TaxID=135249;                                      |   |                 |         |
| OX   | [1]   | RC                        | SEQUENCE FROM N.A.                                      |   |                 |         |
| RC   | STRAIN=AL-210;  | RA                        | Yang S.J., Cha J.;                                      |   |                 |         |
| RT   | "Levan fructotransferase gene (lftm) from Microbacterium sp. AL-210." | RL                        | Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases. |   |                 |         |
| CC   | -/- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.          | DR                        | EMBL; AF294730; AAG01554.1;                             |   |                 |         |
| DR   | InterPro: IPR001362; GH.32.   | DR                        | Pfam: PF00251; Glyco_hydro.32.1.                        |   |                 |         |
| KW   | Glycosidase; Hydrolase; Transferase.                                  | SK                        | SEQUENCE 530 AA; 57458 MW; 1A7E78D19E1D0596             | CRC64;                                  |                 |         |
| Query Match 82.1%; Score 2335; DB 2; Length 530;             |   |                           |   |   |                 |         |
| Best Local Similarity 79.9%; Pred. No. 7e-146;               |   |                           |   |   |                 |         |
| Matches 422; Conservative 36; Mismatches 58; Indels 12; Gaps |   |                           |   |   |                 |         |
| Qy   | 4   | AISRRVVLQAGAGALALIFG----- | -GAVPPAARASAPGSLURAYVHTPPSGWLC                          | 53                                      |                 |         |
| Db   | 5   | SFRRRTMLKGTSGVAGALALITG   | SSPTLQAQSSAGLRAPAAAS--GSLURAYVHTPPSGWLC                 | 62                                      |                 |         |
| Qy   | 54  | DPQRPVTHCAYQLYLIHSD       | QNNPGGWDHASTDCGVAETHHCTVMPLRPDPFVMSGSAV                 | 113                                     |                 |         |
| Db   | 63  | DPQRPVRTGAYQLYLIHSD       | QNNPGGWDHATTSDCGVAETHHCTVMPLRPDPFVMSGSAV                | 122                                     |                 |         |
| Qy   | 114   | VGTTANTAGFCAGAV           | VALATOPTDGVKRYQEOYLYWSTDGGFTTALPDQPVIVNTDGRAAT          | 173                                     |                 |         |
| Db   | 123   | VDTTANTAGFCAGAV           | IALATOPTDGIKRYQEOYLYWSTDGGYFTTALPDQPVILNTDGRAT          | 182                                     |                 |         |
| Qy   | 174   | TPAEIENAEWFRDPK           | IHWDTARGEWCVICIGRLRYAAFYTSPNLRDWTLRNEDYPNHALG           | 233                                     |                 |         |
| Db   | 183   | TPAEIENAEWFRDPK           | IHWDAALRGWCVICIGRYASYFTSTNLRDQWTSNEDYPNHALG             | 242                                     |                 |         |
| Qy   | 234   | GIECPDLFEITADG            | TRHVVLAASMDAYGIGLPMYAYTGTWDEQFIADHDLTQWLQW              | 293                                     |                 |         |
| Db   | 243   | GIECPDLFEITADG            | TRHVVLAASMDAYGVGLPMTYAYTGTWNGTHFTAGNLVFPQWLQW           | 302                                     |                 |         |



```
QY 294 GMDWYAAVTPSIDAPETKRLAIAMNNMKYAAARDVPTDASDGYNGQNSIVRELRLARQP 353
DB 303 GMDWYAAVTPKAEAPETRLAIAMNNMKYAAARDVPTDASDGYNGQNSITRELRLRQE 362
QY 354 GGWYTLTSPVAALTYVATTTLPRTVDGSAVLWPNGRAYEIELDIAWDATNVCISV 413
DB 363 GGWYSLSEFVGALGYVATTTIPTQVNGSVVLWPNGRAYEIELDISWTDAAVGLSV 422
QY 414 GRSPDCTRTNIGKYGADLYVDRGSDLAGYSLAPYSRAAAPIDPCGARSVHLRLIYLDQTS 473
DB 423 GRSADGARTNIGKYGSDLYVDRGSDQAGFALSPTRAAAPIDAGVRSVHLRFVVDQTS 482
QY 474 VEVFNAGHTVLSQQVHFAEGDTGISLYTDGGPAHFTGVIVREIGQAI 521
DB 483 VEVFNAGHTVLSQQVHFAEGDTGISLYSDGGAATSGIRIEFGTPI 530
RESULT 4
ID Q93R69 PRELIMINARY; PRT; 620 AA.
AC Q93R69;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Levanase.
OS Microbacterium laevaniformans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Microbacteriaceae; Microbacterium.
OX NCBI_TaxID=36807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15953;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porvolik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunz F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1;
DR EMBL; Z99121; CAB15451.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Glycosidase; Hydrolase.
SQ SEQUENCE 620 AA; 67667 MW; 82B4909328CBAAC2 CRC64;
Query Match 36.8%; Score 1047.5; DB 2; Length 620;
Best Local Similarity 45.3%; Pred. No. 5.5e-61;
Matches 243; Conservative 55; Mismatches 190; Indels 49; Gaps 16;
QY 3 PAI--SRAVLOG---AGACALALIFGGAVPPAARASAPGSLRAVYHMTPTSCWLCDPOR 57
DB 34 PALPARRALAGLALAGSAAVLAVG---PAAAEEDAP-SURARYHFTVPDHMMNDPQR 89
QY 58 PVTTHGAYQLYLIH-SDQNGPGG-----WDHASTTDGVAFTHHGTVMF--LRPDPVVM 109
DB 90 PVVIDGLVHYLYNADYDANFRANFTGEMRLATSADGVAADOGVAAPKGTNANYDLWS 149
QY 110 GSAVGTANTAGGAGAVVALATQ---PRDGVV---KYQEQYLYNSTDGGTFTALPDP 162
DB 150 GSAVVDHAGTAGGPGAVVMLVTQMDHPTAAQKLDASGQQAFLWYSVDGGRFTFRPDGDE 209
QY 163 VIVNTDGRAATTPAETIENAEWFRDPKHWDRTARGWVVCVIGRLRYAAFTYSPNLRDWTLR 222
DB 210 PVIQDGRD-----FRDPKVWDDERQWRVVALIAERDRVSFTYSPDLHRWT-- 256
QY 223 RNFDYPNHALGGTECPDLFEITADGCTHRHWLAASMDAYGIGLPMTYAYWTGTWDEQFH 282
DB 257 RTGEYVNAIGITIECPDLFRLRADDGTHWVMGVSANGATNEPATYAYWTGSGDSTFV 316
QY 283 ADDLTPQWLDGWMDWYAAVTPSIDAPETKRLAIAMNNMKYAAARDVPTDASDGYNGQNS 342
DB 317 PDVAAPQWLDHGDWYGAFTWEDPAAPLRLAIAMNNNDY-AHGAPTWPDRDGTGTD 375
QY 343 IVRELRLARQPGWYTLTSTPVAALTYVATTTLPRTVDGSAVLWPNGRAYEIELDIA 402
DB 376 ITRIELRLAR-AGAGYSLSPAPDALQDHATHHTALGDVVRVDGFTELAYRGDAYEITRIS 434
```

```
QY 403 WDTATNNGISVGRSPDCTRTNIGKYGADLYVDRGSDLAGYSLAPYS---RAAAPIDPG 459
DB 435 RQDADNVGQLRSADSGSHRADGLTRDYALNRAQTG-----RDSKWKVSERTPLD-D 487
QY 460 ARSVHLRLIYLDQTSVEVFNAGHTVLSQQVHFAEGDTGISLYTDGGPAHFTGVIVRE 516
DB 488 ADTVLRLIYLDQTSVEVFNAGHTVLSQQVHFAEGDTGISLYTDGGPAHFTGVIVRE 544
RESULT 5
ID O07003 PRELIMINARY; PRT; 516 AA.
AC O07003;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein yveb.
GN YVEB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porvolik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunz F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; Z94043; CAB08014.1;
DR EMBL; Z99121; CAB15451.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR Glycosidase; Hydrolase; Hypothetical protein; Complete proteome.
KW
```

|                                 |   |   |
|---------------------------------|---|---|
|                                 | Query Match   | 30.9%; Score 880.5; DB 16; Length 514;  |
|                                 | Best Local Similarity   | 40.7%; Pred. No. 4.3e-50;   |
|                                 | Matches   | 200; Conservative 76; Mismatches 179; Indels 37; Gaps                           |
| Qy                              | 40 RAVYHMTPPSOWLCDPQPVTTHCAYOLYLHS-DQNNGPG-GWDHASTTGDGVAFTHGT 97      | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 45 REVYHTVPNKWKNDPQRPVFFDGEYHYLYLNKDYPNGNGTEMRQATSKDLTWKDEGV 104      | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 98 VMP--LRPDFVMSSGSAVVGTANTAGFGAGAVVALAQTPTDGRVKRQEQLYLVSTDGGFT 155   | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 105 CIPKYTNENGDISSGFVDAQNTAGFKGAIVAVITQPSASMDK-QEQLFWLSTDRGT 163      | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 156 FTALPD-PVIVNTDGRAATTPAEIENAEWFRDPKIHWDTARGECVCVIGRLRYAAFYTS 214   | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 164 FKPYSNQPIIPN-----PCTKD----FRDPKIIWDFKNKKWMVLAEGTKIGFYESY 211      | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 215 NLRDWTLRNFEDYPNHALGIECPDLFEIYADDTGRHWVLAAASMDAYGIGLPMTYAYWTG 274  | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 212 NLKNW--QHTGDFTFNIGIVECPDIFPMOSNDGNCKWILGSTANGKVSGEPNTYAYWG 269    | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 275 TWDCGEOPHADDLTPQWLDKGWDAVAATWFSIDAPET--KRLALAAMNNKKYAAAROVPTD 332 | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 270 NYDGKKTADIISTPKWLDYGFDTAAATFFSEKQSLMKLKRYALAMNNNDYI-ANNFTPT- 327  | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 333 ASDGYNGNSIVRELRLARPGGWYTLLSPVAALTNYVTATTTLDPRTVDGSAVLPFWNG 392    | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 328 IQNFENGMSIVREITLSKQKNTYSLLSPIKRNIETTSDQFKOISVYKGLKHLKVQG 387      | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 393 RAYEIELDIAWDTATNVGISVGRSPDGRTRHTNIGKY--GADLYVDRGPSDLAGYSLAP-- 448 | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 388 KAYQLDITINWVDAKNVGIRLESRLDKRHIDVGIFTPEGKYSVNR-----AYTGPNPK 441    | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 449 ---YSRAAAPIDPCARSVHLRIILLVDTQSVEFVNAGHTVLVSQQVHAEGDTGSLYTDGG 505  | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 442 SKKYVESRAPDDINKKHVLIKFVDKYSVEVFIDDGKITYSNEVFPREDKGITLFSIKG 501    | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Qy                              | 506 PAHFTGTWIREI 517  | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| Dd                              | 502 KAVEKNITIKHI 513  | I : I I I I I : I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I : |
| <br>RESULT 7                    |   |   |
| P94469                          | ID  | PRELIMINARY; PRT; 395 AA.   |
| AC                              | A94469  |   |
| AC                              | A94469;   |   |
| DT                              | 01-MAY-1997 (TrEMBLrel. 03, Created)                                  |   |
| DT                              | 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)                     |   |
| DT                              | 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)                   |   |
| DE                              | Levanase precursor (EC 3.2.1.65) (2,6-beta-D-fructan                  |   |
| DE                              | fructanohydrolase).   |   |
| DE                              | SURC.   |   |
| ON                              | Bacillus stearothermophilus.  |   |
| OS                              | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;         |   |
| OC                              | Geobacillus.  |   |
| NCBI_TaxId=1422;                | NCBI_TaxId=1422;  |   |
| RX                              | [1]   |   |
| RP                              | SEQUENCE FROM N.A.  |   |
| RC                              | STRAIN-ATCC 12980;  |   |
| RX                              | MEDLINE=98007870; PubMed=9349714;                                     |   |
| Li Y., Tricca J.A., Ferenci T.; |   |   |
| RT                              | "A novel levansucrase-levanase gene cluster in Bacillus               |   |
| RT                              | stearothermophilus ATCC12980."  |   |
| RL                              | Biochim. Biophys. Acta 1353:203-208(1997).                            |   |
| CC                              | -  - CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 2,6-BETA-D-FRUCTO       |   |
| CC                              | FURANOSIDIC LINKAGES IN 2,6-BETA-D-FRUCTANS (LEVANS) CONTAINING       |   |
| CC                              | MORE THEN 3 FRUCTOSE UNITS.   |   |
| CC                              | -  - INDUCTION: IN CONTRAST WITH LEVANSUCRASE AND SUCRASE, LEVANASE   |   |
| CC                              | SYNTHESIS IS NOT INDUCIBLE BY SUCROSE; NO INDUCER IS KNOWN FOR        |   |
| CC                              | LEVANASE.   |   |
| CC                              | -  - SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.         |   |
| DR                              | EMBL; U34875; AAB97112.1; --  |   |
| DR                              | InterPro; IPR001362; GH_32.   |   |
| DR                              | Pfam; PF00251; Glyco_hydro_32; 1.                                     |   |

|   |   |  |                                 |      |         |
|---|---|--|---------------------------------|------|---------|
| Db  | 73  | PPAIPGGEDGAIFSGSAVVDKNNSTSGFDEBEGSG-LVAIYTNENKSKP----          | GKPOVQS                         | 127  |         |
| Qy  | 146   | LYWSTDDGGFTTALP-DPVIIVTDDGRAATTPAETIENAEWFRDPKIHWDTARGEWVCVIGR | 204                             |      |         |
| Db  | 128   | IAYSKDGRVWTKYEGNPVIFPTDTLD-----                                | FRDPKVIWHDSSKWIWVLAV            | 175  |         |
| Qy  | 205   | LYRAAFYTSNLRDWTLLRRNF--DYPNHLAGGIECPOLFETIADD--GTRHVLAAASM--   | 258                             |      |         |
| Db  | 176   | RDRVEFTSPNLKESFASFEFGSDIPHHRIGIEPCDIFRIQVDEDLNTTKWILMLSVGD     | 235                             |      |         |
| Qy  | 259   | -----DAYGIGLPMTYAYTWTGDEGFHADDL-----                           | TPOWLWDGWDTAAVT                 | 302  |         |
| Db  | 236   | RNGVNPDPDPPEPPAGSGM---MYFVSGFCKSFPTDEALESIDTIKWVDYGSDFYAAVS    | 291                             |      |         |
| Qy  | 303   | WPSIDAPETKRLATAWNKNKYAARDVPTDASDCYNGNSIIVRELRLAROPGGWYTLTST    | 362                             |      |         |
| Db  | 292   | WNGISNEDGRKIWLGNWNWRYAT----                                    | TLPSEKWRGKTSIPRELQRTYPEG-LRLIQT | 346  |         |
| Qy  | 363   | PVAALTYNVYATTTPLDPTV-DGSAVL-PWNGRAYETIELDIAMDATNVGISVGRSPDGT   | 420                             |      |         |
| Db  | 347   | PINELSQLRPILSLQDLTIKPGMNVLSDISAAKAEIIAEFEIGTAVFEFGFKVRKS--AN   | 404                             |      |         |
| Qy  | 421   | RHTWIGK--YGADLYVDRGSPDLACYSLAPYSRAAIPDPGARSVHLRIILVDTQSVFV     | 478                             |      |         |
| Db  | 405   | QETIIIGNISNEELFVDRTKSSATDFHSDFTAIHKATMKPEHERIQLSIYLDWSSVEVFG   | 464                             |      |         |
| Qy  | 479   | NAGHTVLSOOVHFAEGDTGSLYTDGGPAHFTGVVREI                          | 517                             |      |         |
| Db  | 465   | NHGKAIISDMIFDPFESKGLLEYALGCELRVVSLQIINDL                       | 503                             |      |         |
| RESULT 9  |   |  |                                 |      |         |
| Q9LAL1  | ID  | Q9LAL1   | PRELIMINARY;                    | PRT; | 501 AA. |
| AC  | Q9LAL1  |  |                                 |      |         |
| DT  | 01-OCT-2000   | (TEMBLrel. 15, Created)  |                                 |      |         |
| DT  | 01-OCT-2000   | (TEMBLrel. 15, Last sequence update)                           |                                 |      |         |
| DT  | 01-MAR-2002   | (TEMBLrel. 20, Last annotation update)                         |                                 |      |         |
| GN  | Exo-inulinase.  |  |                                 |      |         |
| DE  | INU2.   |  |                                 |      |         |
| OS  | Pseudomonas mucidolens.   |  |                                 |      |         |
| OC  | Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;        |  |                                 |      |         |
| OC  | Pseudomonas.  |  |                                 |      |         |
| OX  | NCBI_TaxID=46679;   |  |                                 |      |         |
| ON  | [1].  |  |                                 |      |         |
| RP  | SEQUENCE FROM N.A.  |  |                                 |      |         |
| RA  | Kwon Y.-M., Choi Y.-J.;   |  |                                 |      |         |
| RT  | "DNA sequences and expression in Escherichia coli of an exo-inulinase |  |                                 |      |         |
| RT  | gene (inu2) from Pseudomonas mucidolens."                             |  |                                 |      |         |
| RL  | Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.               |  |                                 |      |         |
| CC  | -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.          |  |                                 |      |         |
| DR  | EMBL: AF129819; AAF41125.1;   |  |                                 |      |         |
| DR  | InterPro: IPR001362; GH_32.   |  |                                 |      |         |
| DR  | Pfam: PF00251; Glyco_Hydro_32.1.                                      |  |                                 |      |         |
| DR  | PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.                             |  |                                 |      |         |
| KW  | Glycosidase; Hydrolase.   |  |                                 |      |         |
| SQ  | SEQUENCE 501 AA; 57797 MW; 4043B86F2084A4E7 CRC64;                    |  |                                 |      |         |
| Query Match   |   |  |                                 |      |         |
| Best Local Similarity 31.3%; Pred. No. 7.4e-27;               |   |  |                                 |      |         |
| Matches 156; Conservative 72; Mismatches 223; Indels 47; Gaps |   |  |                                 |      |         |
| Qy  | 40  | RAVYHMTPPSGWLCDPQRPVTHGAYOLYLHSDNN--                           | GPGGWDHASTTDDGVAFTHHGT          | 97   |         |
| Db  | 11  | RPOHFHTPETNMNDPNGLVYEGEYHLFTQYHPFGTGMHGWCHAVSKDLIHWDH--L       | 68                              |      |         |
| Qy  | 98  | VMPLRPDF--PVWSGSAVGTANTAGF--GAGAVVALATQP---                    | TDGVRKYQEQLYWS                  | 150  |         |
| Db  | 69  | PIALYPDHGQIFSGSAVDQWDTSGFNGGAGWALFTHADKFPDPSGRPRQSLAYST        | 128                             |      |         |
| Qy  | 151   | DGGFTFTALDPVIVTDDGRAATTPAETIENAEWFRDPKIHWDTARGEWVCVIGRLYAAF    | 210                             |      |         |

```

Db 129 DRGRW-----IKYGNPFDAEAMTD---FRDPKVFVHKASNKWIMILAGNCVRI 177
QY 211 YTSPLRDLWTLRRNPDYFNHALGGI-ECPDLFETADGTRH---WVLAASM---DAYGI 263
Db 178 YHSSNLKQWKASERGANEGSHGTGWECPLDFELPV-DGDRNRKKVWVVSIGNSDEYLE 236
QY 264 GLPMTYAYWTGTWGEQF--HADDLTJQWLDGMDWYAAVTPSIDAPETKRLAIAMNN 321
Db 237 G--SRTQYFIGQFDGIRFTENDPETYLVWDHGRDNYAGVTSWDIPEEDGRRLFIAMSN 294
QY 322 WKYAADVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALTYNYVATTPLDR- 380
Db 295 WKY-ANHTPTET---WRSAMTIPKSLSRSTPEG-TRLFPVPTVELESRLKLEWKDLE 349
QY 381 -TVDSAVLPWNGRAYETELDIAMDATNVGISVGRSPDGRHNTNIGYKAD---LYVDR 436
Db 350 VTTNGVNLNSIGDIFEIIAEFELNTALEFGFKVRKSD--QETIVG-YDAEQQMLFIDR 406
QY 437 GPSDLAGYSLAPYRAAAPIPDGARSVHLRLVDTQSVVEFVNAGHTVLSOOVHFAEGDT 496
Db 407 TQSGVSDCEHFPCPKHGAVMIPNQRIQMHIFVDRSSVEFGNGELTMTDLIPDDSS 466
QY 497 GISLYTDCGPAHFTCIIV 514
Db 467 GIEVTAKEGNVKLVSML 484

RESULT 10
Q45155 ID Q45155 PRELIMINARY; PRT; 622 AA.
AC Q45155;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Levanase.
GN SCRL.
OS Bacteroides fragilis.
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BF1;
RX MEDLINE=93259952; PubMed=8491724;
RA Blatch G.L., Woods D.R.;
RT "Molecular characterization of a fructanase produced by Bacteroides fragilis BF-1."
RL J. Bacteriol. 175:3058-3066(1993).
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; M83774; AAA2924.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 622 AA; 70287 MW; 63424EB274CEE134 CRC64;

Query Match 18.18; Score 515; DB 2; Length 622;
Best Local Similarity 30.28; Pred. No. 6.6e-26;
Matches 156; Conservative 77; Mismatches 195; Indels 88; Gaps 20;

QY 40 RAVYHMTPPSCWGLCDPQRPVTHTCAYQLYLHSDQNN--GPGGWDHASTTDCGVAFTHGT 97
Db 133 RPLVHHTPLYGWMDANGLVTKGDEYHLFYQYNPYGSWMGNHWHGSVKDLVHWEH--- 189
QY 98 VWLPRPOFP-----VMSGSVAVGTATAGGAVVALATOPTDGVGRKYEQVLYWSTD 151
Db 190 ---LEPALARDTLGHIHSGSVDDANTAGYAGAIYFYSASD--KNGOIQCMAYSTD 244
QY 152 GGFTFTAL-PDPVIVNTDGRAATTPABIENAEWFRDPKIHWDTARGEMVCVIGRLRYAAF 210
Db 245 NGRFTTYEKNPVLTPTD-----LKDRPQKVEWYAPDQKWMVVSADKEMRF 293
QY 211 YTSPLRDLWTLRRNFD-----YPNHALGGITCPLDFELTADGTRH---WVLAASMDAYGI 263

```

```

Db 294 YSENKLEWTVMSGWEGYGVQPSQ-----FECPDWVELPVDGNPDHKKWALIVNN----- 345
QY 264 GLPMTY-----AYWTGTWGEQFHADDL--TPQWLDGMDWYAAVTPSIDAPETKRLA 315
Db 346 --PCYFGGSATQYFIGDFGEKFCVCDNKPETVWLDGMDKHATYATVCFESNTG---DRTIA 400
QY 316 IAMNNKYYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYLLSTPVAALTYNYVATT 375
Db 401 VPMNSNQY-ANIVPTQ---FRSANALPRELSLYTQDGDYI-NAAPVVEETKSLRKESR 455
QY 376 TLPDRTVDGS-----AVLPWNGRAYETELDIAMDATNVGISVGRSPDGRHNTNIGYKAD 431
Db 456 EIPAFEGVDAYHVDLSNDKNGAYEIELELAAGSAEIMGLKLFNEKGENVDYIYLSPEKK 515
QY 432 LYVDRGPSDLA--GYSLAPYSRAA-----APIDPGARSVHLRLTV 469
Db 516 LVMDRTKSGIVDFGKDSAPHAEIHDRRKQNSINYVDFAELGTWAPVQK-AGNYKLDIFV 574
QY 470 DTQSVVEFVNAGHTVLSQOVHFAEGDTGISLYTDCG 505
Db 575 DKCSVEIFLNGKKTAMTNLIPPTTPYQMSFYSRGG 610

RESULT 11
Q9RBJ1 ID Q9RBJ1 PRELIMINARY; PRT; 534 AA.
AC Q9RBJ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Levanase precursor.
GN LSDB.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=33996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SRT4;
RX MEDLINE=96253999; PubMed=8704949;
RA Arrieta J., Hernandez L., Coego A., Suarez V., Balmori P.,
Menendez C., Petit-Glatron M.F., Chambert R., Selman-Housein G.;
RT "Molecular characterization of the levansucrase gene from the
endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4."
RL Microbiology 142:1077-1085(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SRT4;
RX MEDLINE=99373266; PubMed=10441728;
RA Hernandez L., Arrieta J., Betancourt L., Falcon V., Madrazo J.,
Coego A., Menendez C.;
RT "Levansucrase from Acetobacter diazotrophicus SRT4 is secreted via
periplasm by a signal-peptide-dependent pathway."
RL Curr. Microbiol. 39:146-152(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SRT4;
RA Menendez C., Hernandez L., Mendoza M.F., Hevia P., Selman G.,
Arrieta J.;
RT "Molecular cloning and expression in E. coli of an exolevanase from
Acetobacter diazotrophicus SRT4."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; L41732; AAF16405.1;
DR InterPro; IPR001362; GH_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 534 LEVANASE.
SQ SEQUENCE 534 AA; 58367 MW; 7BDBA4A9BA478F6 CRC64;

Query Match 18.08; Score 513; DB 2; Length 534;

```

|                       |   |  |                    |                |                        |                    |               |     |
|-----------------------|---|--|--------------------|----------------|------------------------|--------------------|---------------|-----|
| Qy                    | 98  | VMPLRPD--FPVWGS                        | SAVGTA             | TANTAGE--GAGAV | LATO-----PTDGV         | KYQEOYLW           | 148           |     |
| Dq                    |   | : : :                                  | : : : : : :        | : : : :        | : : : :                | : : : : : :        |               |     |
| Db                    | 68  | P1ALEPDELGT                            | IFSGSAVD           | EQDTSGFPGK     | GPLVAIFTHHSL           | PLGPTEQVR--QCOSL   | 125           |     |
| Qy                    | 149   | STDCGGFTALP                            | -DPVLVTND          | GRAATTPAETI    | ENAEFRDKIHWD           | TARGEWCVIGRLRY     | 207           |     |
| Db                    |   | : : : : : :                            | : : : :            | : : : : : :    | : : : : : :            | : : : : : :        |               |     |
| Db                    | 126   | STDGSKTWIKT                            | GNPVL-----         | EDKECIDR       | PKPVFWHKPTQ            | WNVMVLACGQT        | 173           |     |
| Qy                    | 208   | AAFYTSPNLRD                            | WTLRRNF            | DYPNALGG---IEC | POLFEITADDDGRH--WVLAAS | M-DA               | 260           |     |
| Db                    |   | : : : : : :                            | : : : : : :        | : : : : : :    | : : : : : :            | : : : : : :        |               |     |
| Db                    | 174   | IRIVHSNPKLE                            | WTFASEF---GHG      | IGSHDAWVECP    | DLFPPLYIDEKRE          | QVKVMVLVSIGDV      | 230           |     |
| Qy                    | 261   | YGIGLPMTYYA                            | NWGTWDGE           | FQHADDLTP--QWL | DWGSDNYAAVTW           | SIDAPETRKLIAIW     | 318           |     |
| Db                    |   | : : : : : :                            | : : : : : :        | : : : : : :    | : : : : : :            | : : : : : :        |               |     |
| Db                    | 231   | PEIREGSRKT                             | OFTTFEGFGT         | TVADAASEKV     | RWLDYGRDNVAG           | CWSDDIPAE          | DGRRLLFMGW    | 290 |
| Qy                    | 319   | MNNKYAARDV                             | PTDASDGY           | NGQNSIVREL     | RLARQP                 | GGWYTLSTP          | VAALTNYVATTTL | 378 |
| Db                    |   | : : : : : :                            | : : : : : :        | : : : : : :    | : : : : : :            | : : : : : :        | : : : : : :   |     |
| Db                    | 291   | MSNWRY-ANQ                             | TPTR---WRG         | AMSIPRELAET    | RK-GTVALVQR            | PVRELEG            | LTPVLSLT      | 345 |
| Qy                    | 379   | DRTVD--GS                              | AVLPWNGRAYE        | FIELDIAWDTAT   | NVGISVGRSP             | DGTRHNICKYGA--DLXV | 434           |     |
| Db                    |   | : : : :                                | : : : : : :        | : : : : : :    | : : : : : :            | : : : : : :        | : : : : : :   |     |
| Db                    | 346   | EPSWEVERNALS                           | ALQDCLYE           | VAEFA--TTGD    | EFGKRVRSDE--QETLV      | GYTSATOEVI         | 401           |     |
| Qy                    | 435   | DRGPSDLAGY                             | --SLAPS            | YRAAAPIDGARS   | YHLRIILVD              | TQSVEVFVNAGHT      | VLSQQVHFA     | 492 |
| Db                    |   | : : : :                                | : : : : : :        | : : : : : :    | : : : : : :            | : : : : : :        | : : : : : :   |     |
| Db                    | 402   | DRKRSGRSD                              | HEDFAGKHG          | ASLTILD--RMD   | IIRYVDRSSIE            | VFDFRGQVAIT        | DLI-FP        | 458 |
| Qy                    | 493   | EGDT-GIS                               | LYT                | 502            |                        |                    |               |     |
| Db                    |   | : : : :                                | : : : : : :        | : : : : : :    | : : : : : :            | : : : : : :        | : : : : : :   |     |
| Db                    | 459   | DAEAKGLEIFS                            | 469                |                |                        |                    |               |     |
| RESULT                | 13  |  |                    |                |                        |                    |               |     |
| Q9APY5                | ID  | Q9APY5                                 | PRELIMINARY;       | PRT;           | 493                    | AA.                |               |     |
| AC                    | OT  | Q9APY5                                 |                    |                |                        |                    |               |     |
| DT                    | 01-JUN-2001   | (TREMBrel. 17, Created)                |                    |                |                        |                    |               |     |
| DT                    | 01-JUN-2001   | (TREMBrel. 17, Last sequence update)   |                    |                |                        |                    |               |     |
| DT                    | 01-MAR-2002   | (TREMBrel. 20, Last annotation update) |                    |                |                        |                    |               |     |
| DE                    | Exo   | nulinase (EC 3.2.1.80).                |                    |                |                        |                    |               |     |
| DS                    | Bacillus subtilis.  |  |                    |                |                        |                    |               |     |
| OC                    | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;         |  |                    |                |                        |                    |               |     |
| OC                    | Bacillaceae; Bacillus.  |  |                    |                |                        |                    |               |     |
| OX                    | NCBI_TaxID=1423;  |  |                    |                |                        |                    |               |     |
| RN                    | [1]   |  |                    |                |                        |                    |               |     |
| RP                    | SEQUENCE FROM N.A.  |  |                    |                |                        |                    |               |     |
| RC                    | STRAIN-SNU-7;   |  |                    |                |                        |                    |               |     |
| RA                    | Koo B.-S., Kang S.-I., Kim S.-I.;                                     |  |                    |                |                        |                    |               |     |
| RT                    | "Nucleotide sequencing of the gene for exoinulinase from Bacillus sp. |  |                    |                |                        |                    |               |     |
| RT                    | snu-7 and its expression in Escherichia coli.";                       |  |                    |                |                        |                    |               |     |
| RL                    | Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.              |  |                    |                |                        |                    |               |     |
| CC                    | CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.       |  |                    |                |                        |                    |               |     |
| DR                    | EMBL: AF244992; AAK00768.1;   |  |                    |                |                        |                    |               |     |
| DR                    | InterPro: IPR001362; GH_32.   |  |                    |                |                        |                    |               |     |
| DR                    | Pfam: PF00251; Glyco_Hydro_32; 1.                                     |  |                    |                |                        |                    |               |     |
| KW                    | Glycosidase; Hydrolase.   |  |                    |                |                        |                    |               |     |
| SQ                    | SEQUENCE 493 AA; 53604 MW; 61E44A79A1600447 CRC64;                    |  |                    |                |                        |                    |               |     |
| Query Match           |   | 16.5%;                                 | Score 469.5;       | DB 2;          | Length 493;            |                    |               |     |
| Best Local Similarity |   | 29.4%;                                 | Pred. No. 4.8e+23; |                |                        |                    |               |     |
| Matches 148;          | Conservative  | 70;                                    | Mismatches 212;    | Indels         | 73;                    | Gaps               |               |     |
| Qy                    | 15  | GAGALALIFG                             | AVPPAARAS          | AGSLRAVYHMT    | PPSGWLCDPQ             | RPVTTHGAYQLYLHSDQ  | 74            |     |

|    |     |   |     |
|----|-----|---|-----|
| Db | 17  | GYGKLTREL--GTAASLAAATLVG--RPILHYTAKNTWLNDPNCGLVQHWGQVHYHLUFYNNPFF | 72  |
| QY | 75  | NN--GGPGWDHASTDGVAFTHHGTVMPLRDPFVWSSGSAWGTANTAGFAGAVVALAT         | 132 |
| Db | 73  | DNWGNNSWGHATSDLLHWHPEVALACNEEDDFVSGTIVDGHGNTSGFGTAEDPALVA         | 132 |
| QY | 133 | QPTDGVARK-----YQEQYLWSTDGGFTTALTALPDPIVNTDGRAATTFAETENAEWFRD      | 186 |

Db 133 IYTSFAKESVHQGTQOASLAFSTDAAGMTWS-----KYAGNPVLGRGSAHFRD 180  
 Qy 187 PKI--HMDTARGEVWCVIGRLRY--AAFTYSPNLRDWTLRNEDYPNHALGGIECPDLFE 242  
 Db 181 PKVRYEGPAGSRIMWAVEAQHQVVLYRSADLKDWEYLSTFGPANASDGEWECPP--LP 238  
 Qy 243 ITADGGRHWVLAASMDAYGIGLPMYVYWTGTWDSQFHA--DDLTPQ----- 289  
 Db 239 VDGPDNVKWLVLVNPNGAVAGSGGQYFVGDFGVQFPADPSLPTDADGNVDLRHC 298  
 Qy 290 -WLDGWDWYAAYTWPISIDAPETKRLAIAWMNNKYYAARDVPTDASDGYNGQNSIVREL 348  
 Db 299 LWLDGWDYAAVSFS--NAPENRRIMIGWNNWY--ANSLPTSP--WRSSMSLAREIE 352  
 Qy 349 LARQPGGWYTLSTPVAALTYVATT-----TLPDRTVDCSVAFLPWNGRAYEI 397  
 Db 353 LA--TVDGFPRLVQRPVPLDSGEPARTIONMELTTPCCNCPTRRPPGSAQL-----I 403  
 Qy 398 ELDIAMDTATNVGIVSGRSPDGRHT--NIGKYGADLYVDRGPSDLAGYSLAPYSRAAAP 455  
 Db 404 EAEILPCTARHIAFRLLGAPDGSAAATVLSFDALTSRLTLDRNSGNTAFTISSRLNPQ 463  
 Qy 456 IDPG-ARSVHLRLVDTQSVVEF 477  
 Db 464 VKASTGGALRLKVIIDQCSVEF 486

## RESULT 14

Q93T55 ID Q93T55 PRELIMINARY; PRT; 493 AA.  
 AC Q93T55;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Exonulinase (EC 3.2.1.80).  
 GN EXO1.  
 OS Bacillus sp. snu-7.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=159251;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SNU-7;  
 RA Koo B.-S., Pan C.-H., Kim S.-I.;  
 RT "Molecular Cloning of Exonulinase Gene from Bacillus sp. snu-7 and  
 its Expression in *Escherichia coli*,"  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
 DR EMBL: AF366292; AAK54126.1;  
 DR InterPro: IPR001362; GH\_32.  
 DR Pfam: PF00251; Glyco\_Hydro\_32; 1.  
 KW Glycosidase; Hydrolase.  
 SQ SEQUENCE 493 AA; 53544 MW; ECEB9C9976BD0E46 CRC64;

Query Match 16.5%; Score 468.5; DB 2; Length 493;  
 Best Local Similarity 29.4%; Pred. No. 5.6e-23;  
 Matches 148; Conservative 70; Mismatches 212; Indels 73; Gaps 18;

Qy 15 GAGALALIFGAVPPPARASAPGSLRAVYHMTPPSGHLCDPQRPVTHGAYQLYLSHQ 74  
 Db 17 GYGKLTSL--GTAASLAAATVGG--RPLHYTAKNTLNDPGLVHWQGVYHLFYQNPPF 72  
 Qy 75 NN--GPGGWDHASTDGVAFTHGTVPLRPDPFVWSGSAVVGVTANTAGFCAGAVVALAT 132  
 Db 73 DNVWGNWSGHAATSTDLHHTHEHPVATACNEEDVFSGSIVDGHGNTSGFGTADPALVA 132  
 Qy 133 OPTDGVK-----YQDYLYWSTDGGFTFTALPDVPVINTDGRAATPAEIEAENAFRD 186  
 Db 133 IYTSFAKESVHQGTQOASLAFSTDAAGMTWS-----KYAGNPVLGPGSAHFRD 180  
 Qy 187 PKI--HMDTARGEVWCVIGRLRY--AAFTYSPNLRDWTLRNEDYPNHALGGIECPDLFE 242

Db 181 PKVRYEGPAGSRIMWAVEAQHQVVLYRSADLKDWEYLSTFGPANASDGEWECPP--LP 238  
 Qy 243 ITADGGRHWVLAASMDAYGIGLPMYVYWTGTWDSQFHA--DDLTPQ----- 289  
 Db 239 VDGPDNVKWLVLVNPNGAVAGSGGQYFVGDFGVQFPADPSLPTDADGNVDLRHC 298  
 Qy 290 -WLDGWDWYAAYTWPISIDAPETKRLAIAWMNNKYYAARDVPTDASDGYNGQNSIVREL 348  
 Db 299 LWLDGWDYAAVSFS--NAPENRRIMIGWNNWY--ANSLPTSP--WRSSMSLAREIE 352  
 Qy 349 LARQPGGWYTLSTPVAALTYVATT-----TLPDRTVDCSVAFLPWNGRAYEI 397  
 Db 353 LA--TVDGFPRLVQRPVPLDSGEPARTIONMELTTPCCNCPTRRPPGSAQL-----I 403  
 Qy 398 ELDIAMDTATNVGIVSGRSPDGRHT--NIGKYGADLYVDRGPSDLAGYSLAPYSRAAAP 455  
 Db 404 EAEILPCTARHIAFRLLGAPDGSAAATVLSFDALTSRLTLDRNSGNTAFTISSRLNPQ 463  
 Qy 456 IDPG-ARSVHLRLVDTQSVVEF 477  
 Db 464 VKASTGGALRLKVIIDQCSVEF 486

## RESULT 15

Q44109 ID Q44109 PRELIMINARY; PRT; 943 AA.  
 AC Q44109;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Levansase.  
 GN LEVJ.  
 OS Actinomycetes naeslundii.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.  
 OX NCBI\_TaxID=1655;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-T14V;  
 RX MEDLINE=95129923; PubMed=7828936;  
 RA Norman J.M., Bunney K.L., Giffard P.M.;  
 RT "Characterization of levJ, a sucrose/fructanase-encoding gene from  
 Actinomycetes naeslundii T14V, and comparison of its product with other  
 sucrose-cleaving enzymes.";  
 RL Gene 152:93-98(1995).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
 DR EMBL: U12274; AAA57876.1;  
 DR InterPro: IPR001362; GH\_32.  
 DR Pfam: PF00251; Glyco\_Hydro\_32; 1.  
 DR PROSITE: PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
 KW Glycosidase; Hydrolase.  
 SQ SEQUENCE 943 AA; 99267 MW; 801A2A7D074579B9 CRC64;

Query Match 16.4%; Score 467.5; DB 2; Length 943;  
 Best Local Similarity 21.5%; Pred. No. 1.6e-22;  
 Matches 187; Conservative 83; Mismatches 217; Indels 381; Gaps 27;

Qy 4 AISRRVLOQAGAGALALIFGG---AVPPAARASAPGS--LRAYVHMTPPSGHLCDPQRP 58  
 Db 19 AASTLASMPAASGQARPVSAAAPAPAPKADQGTGERWRPQSHYTPKQNMNDPGL 78  
 Qy 59 VTHGAYQLYLSHQNN--GPGGWDHASTDGVAFTHGTVPLRPDPFVWSGSAVVGVT 116  
 Db 79 VYDGEYHMFYQYNPEGSDWGNMWSGHAISKDLVHWQELGVAIPHTSQYGVFSGSAVIDT 138  
 Qy 117 ANTAGFGA---GAVVALATQPTDGVKRYQDYLYWSTDGGFTFTAL--PDVPVINTDGRA 171  
 Db 139 KNTSGLGSPDNPAMVAVMTADYGGN--QSQSLAYSTDKGTTWNLNNGDPVL----- 189  
 Qy 172 ATPAEIENAEWFRDPIKHWDTARGEWVCVIGRL--RYAAFTYSPNLRDWTLRNEDYPN 229  
 Db 190 -----DIGSNE-FRDPKPVFDQASGRWTWVVSATHRVFSFYSPDLIHWTEQSGFEGG 243

```
QY 230 HALGIECPDLEITADDGTRH--WVLA----- 255
Db      ||| | : | : |||
244 ITSVMACPDFPEPLVDGSSQEVKVLVTVVADSAQYFVGSWDGTTFTPEIPHYSGEG 303
QY 256 -----ASMDAYG----- 262
Db      | : | : |
304 TTLADFENGYAGWKADGAAGFGSGPATGDLPGHQKAYVDSFGSGDADTGLTSDFTVSS 363
QY 263 ----- 262
Db 364 SYINLRTAGGHPYNPQATGDNNGGRLLAGFDGSWEGWTEGSAFAATPPQCATPAOQPL 423
QY 263 -----IGLPMY-----AY----- 271
Db      || ||
424 VNHSSAGLINTYLDAAATQGGSDAPTGTATSPFTTIDSAYLNLMLGGGNNRPRGGADGGS 483
QY 272 -----W-----TGTW----- 276
Db 484 RVSVELIVDGKVRSRATGRNLEELNMQSDVSDLKGSQAQIVVTDATGGMGHILLDEV 543
QY 277 ----- 276
Db 544 RASDKKASPIADNTSVNLVVDGKVVASATGNNSGTLEWTSMMVAAKGRKARLVIEDRNG 603
QY 277 -----DGEQFHADDLTPQWLDGWDWYAAVTPSIDAPETKRLAIAMNN 320
Db      | : | : | : ||| | : | : |||
604 NAEDWGHLMVDQILQSDTKAFSGADWVPR-LDYKDYAAVTDNV--PNGKRYQVGWMS 660
QY 321 NWKYAARDVPTDASDGYNGQNSIVRELRLARQPGGWYTLSTPVAALTNVTA-----T 374
Db      ||| ||| : ||| : | : | : | : |
661 NWAY-VRDLPTTT---WRTAMSTVREMGLTR-VNGKRLTAQPVTALESRTGOELIRKD 715
QY 375 TPLP-DRTVDGSAVLPMNGRAYEIELDIADWTATNVGISVGRSPDGRHTNIG--KYGAD 431
Db      | : | | : | : | : | : | : | : | : | : |
716 TDIPVGETSLGKAA---QGTSLDISVDLSPSASSFAGLKV--LDNGEQYTLICYDSQAKQ 770
QY 432 LYVDRGPSDLAGYSLAPYSRAAAPIDPGAR-SVHLRIILVDTQSVEFVNAGHTVLSQOVH 490
Db      ||| | : | : | : | : | : | : | : | : | : |
771 LVVDRTHSGVTDSPKFFPARSTAPLSPDSKGQVHLRIIVDAHSVEVFAADGTPVITQVY 830
QY 491 FAEGDTGISLYTDGGPAHFTGIVVREIG 518
Db      : ||| : ||| : ||| : ||| : |||
831 PEQDATGVSLYAEGGTAHLGSLSLHLG 858
```

Search completed: May 6, 2003, 12:34:41  
Job time : 61.1096 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**